Fast strong approximation Monte-Carlo schemes for stochastic volatility models

Christian Kahl* Peter Jäckel†

First version: 28th September 2005
This version: 22nd May 2006

Abstract

Numerical integration methods for stochastic volatility models in financial markets are discussed. We concentrate on two classes of stochastic volatility models where the volatility is either directly given by a mean-reverting CEV process or as a transformed Ornstein-Uhlenbeck process. For the latter, we introduce a new model based on a simple hyperbolic transformation. Various numerical methods for integrating mean-reverting CEV processes are analysed and compared with respect to positivity preservation and efficiency. Moreover, we develop a simple and robust integration scheme for the two-dimensional system using the strong convergence behaviour as an indicator for the approximation quality. This method, which we refer to as the *IJK* (4.47) scheme, is applicable to all types of stochastic volatility models and can be employed as a *drop-in replacement* for the standard log-Euler procedure.

Acknowledgment: The authors thank Vladimir Piterbarg and an anonymous referee for helpful comments and suggestions.

1 Introduction

Numerical integration schemes for differential equations have been around nearly as long as the formalism of calculus itself. In 1768, Euler devised his famous stepping method [Eul68], and this scheme has remained the fallback procedure in many applications where all else fails as well as the benchmark in terms of overall reliability and robustness any new algorithm must compete with. Many schemes have been invented since, and for most engineering purposes involving the numerical integration of ordinary or partial differential equations there are nowadays a variety of approaches available.

With the advent of formal stochastic calculus in the 1920's and the subsequent application to real world problems came the need for numerical integration of dynamical equations subject to an external force of random nature. Again, Euler's method came to the rescue, first suggested in this context by Maruyama [Mar55] whence it is also sometimes referred to as the *Euler-Maruyama* scheme [KP99].

An area where the calculus of stochastic differential equations became particularly popular is the mathematics of financial markets, more specifically the modelling of financial movements for the purpose of pricing and risk-managing derivative contracts.

Most of the early applications of stochastic calculus to finance focussed on approaches that permitted closed form solutions, the most famous example probably being the Nobel prize winning article

*Quantitative Analytics Group, ABN AMRO, 250 Bishopsgate, London EC2M 4AA, UK, and Department of Mathematics, University of Wuppertal, Gaußstraße 20, Wuppertal, D-42119, Germany

†Global head of credit, hybrid, commodity, and inflation derivative analytics, ABN AMRO, 250 Bishopsgate, London EC2M 4AA, UK

by Fischer Black and Myron Scholes [BS73]. With increasing computer power, researchers and practitioners began to explore avenues that necessitated semi-analytical evaluations or even required fully numerical treatment.

A particularly challenging modelling approach involves the coupling of two stochastic differential equations whereby the diffusion term of the first equation is explicitly perturbed by the dynamics of the second equation: stochastic volatility models. These became of interest to financial practitioners when it was realised that in some markets deterministic volatility models do not represent the dynamics sufficiently. Alas, the first publications on stochastic volatility models [Sco87, Wig87, HW88] were ahead of their time: the required computer power to use these models in a simulation framework was simply not available, and analytical solutions could not be found. One of the first articles that provided semi-analytical solutions was published by Stein and Stein [SS91]. An unfortunate feature of that model was that it did not give enough flexibility to represent observable market prices, i.e. it did not provide enough degrees of freedom for calibration. In 1993, Heston [Hes93] published the first model that allowed for a reasonable amount of calibration freedom permitting semi-analytical solutions. Various other stochastic volatility models have been published since, and computer speed has increased significantly. However, despite the fact that at the time of this writing computer power makes fully numerical treatment of stochastic volatility a real possibility, comparatively little research has been done on the subject of efficient methods for the numerical integration of these models. In this article, we present and discuss some techniques that help to make the use of fully numerically integrated stochastic volatility models a viable alternative to semi-analytic solutions, despite the fact that major advances on the efficient implementation of Heston's model have been made [KJ05]. In section 2, we present the specific stochastic volatility models that we subsequently use in our demonstrations of numerical integration methods, and discuss some of their features in the context of financial markets modelling. In section 3, we elaborate on specific methods suitable for the volatility process in isolation. Next, in section 4, we discuss techniques that accelerate the convergence of the numerical integration of the combined system of stochastic volatility and the directly observable financial market variable both with respect to the discretisation refinement required and with respect to CPU time consumed. This is followed by the presentation of numerical results in section 5. Finally, we conclude.

2 Some stochastic volatility models

We consider stochastic volatility models of the form

$$\mathrm{d}S_t \;=\; \mu S_t \mathrm{d}t + V_t^p S_t \mathrm{d}W_t \tag{2.1}$$

where S describes the underlying financially observable variable and V, depending on the coefficient p given by the specific model, represents either instantaneous variance ($p = 1/2$) or instantaneous volatility ($p = 1$).

As for the specific processes for instantaneous variance or volatility, we distinguish two different kinds. The first kind is the supposition of a given stochastic differential equation directly applied to the instantaneous variance process. Since instantaneous variance must never be negative for the underlying financial variable to remain on the real axis, we specifically focus on a process for variance of the form [Cox75, CR76, Bec80, AA00, CKLS92]

$$\mathrm{d}V_t = \kappa(\theta - V_t)\mathrm{d}t + \alpha V_t^q \mathrm{d}Z_t, \qquad V_{t_0} = V_0\ . \tag{2.2}$$

with $\kappa, \theta, \alpha, q > 0$, and $p = 1/2$ in equation (2.1). We assume the driving processes W_t and Z_t to be correlated Brownian motions satisfying $\mathrm{d}W_t \cdot \mathrm{d}Z_t = \rho\,\mathrm{d}t$.

The second kind of stochastic volatility model we consider is given by a deterministic transformation

$$\sigma_t = \sigma_0 \cdot f(y_t)\ , \qquad f : \mathbb{R} \to \mathbb{R}_+\ , \tag{2.3}$$

with $f(\cdot)$ being strictly monotonic and differentiable, of a standard Ornstein-Uhlenbeck process

$$\mathrm{d}y_t = -\kappa y_t \mathrm{d}t + \alpha\sqrt{2\kappa}\,\mathrm{d}Z_t, \qquad y_{t_0} = y_0\ , \tag{2.4}$$

setting $V_t = \sigma_t$ and $p = 1$ in equation (2.1). The transformation $f(\cdot)$ is chosen to ensure that $\sigma \geq 0$ for the following reason. It is, in principle, possible to argue that instantaneous volatility is undefined with respect to its sign. However, when volatility and the process it is driving are correlated, a change of sign in the volatility process implies a sudden change of sign in effective correlation, which in turn implies a reversal of the conditional forward Black implied volatility skew, and the latter is a rather undesirable feature to have for reasons of economic realism. As a consequence of this train of thought, we exclude the Stein & Stein / Schöbel & Zhu model [SS91, SZ99] which is encompassed above by setting $f(y) = y$.

In order to obtain a better understanding of the different ways to simulate the respective stochastic volatility model we first give some analytical properties of the different approaches.

2.1 The mean-reverting CEV process

By *mean-reverting CEV process* we mean the family of processes described by the stochastic differential equation (2.2). Heston's model, for instance, is given by $q = 1/2$ with $p = 1/2$ in the process for the underlying (2.1). The family of processes described by (2.2) has also been used for the modelling of interest rates [CKLS92].

For the special case $q = 1/2$, i.e. for the Heston variance process, the stochastic differential equation is also known as the Cox-Ingersoll-Ross model [CIR85]. In that case, the transition density is known analytically as

$$p(t_0, t, V_{t_0}, V_t) = \chi^2_d\left(\nu V_t, \xi\right) \tag{2.5}$$

with

$$\nu = \frac{4\kappa}{\alpha^2\left(1 - e^{-\kappa\Delta t}\right)} \tag{2.6}$$

$$\xi = \frac{4\kappa e^{-\kappa\Delta t}}{\alpha^2\left(1 - e^{-\kappa\Delta t}\right)} V_{t_0} \tag{2.7}$$

$$\Delta t = t - t_0 \tag{2.8}$$

$$d = \frac{4\theta\kappa}{\alpha^2} \tag{2.9}$$

where $\chi^2_d(x, \xi)$ denotes the noncentral chi-square density of variable x with d degrees of freedom and non-centrality parameter ξ. Broadie and Kaya used this transition density for the Monte-Carlo simulation of European options [BK04].

With $q = 1$, equation (2.2) turns into a stochastic differential equation which is affine in the drift and linear in the diffusion also known as the Brennan-Schwartz model [BS80]. To the best of our knowledge, there are no closed form explicit solutions for this equation allowing for a fully analytical expression, despite its apparent simplicity. A formal solution for equations of the form

$$\mathrm{d}X_t = \left(a_1(t)X_t + a_2(t)\right)\mathrm{d}t + \left(b_1(t)X_t + b_2(t)\right)\mathrm{d}W_t, \qquad X_{t_0} = X_0\ , \tag{2.10}$$

is described in [KP99, Chap. 4.2 eq. (2.9)] as

$$X_t = \Xi_{t_0,t} \cdot \left(X_0 + \int_{t_0}^{t} \frac{a_2(s) - b_1(s)b_2(s)}{\Xi_{t_0,s}} \mathrm{d}s + \int_{t_0}^{t} \frac{b_2(s)}{\Xi_{t_0,s}} \mathrm{d}W_s \right) \tag{2.11}$$

with $\Xi_{t_0,t}$ given by [KP99, Chap. 4.2 eq. (2.7)]

$$\Xi_{t_0,t} = \mathrm{e}^{\int_{t_0}^{t}\left(a_1(s) - \frac{1}{2}b_1^2(s)\right)\mathrm{d}s + \int_{t_0}^{t} b_1(s)\mathrm{d}W_s}\ . \tag{2.12}$$

Applying this to equation (2.2) with $a_1(t) = -\kappa$, $a_2(t) = \kappa\theta$, $b_1(t) = \alpha$ and $b_2(t) = 0$ leads to

$$\Xi_{t_0,t} = \mathrm{e}^{-\left(\kappa+\frac{1}{2}\alpha^2\right)(t-t_0)+\alpha(W_t-W_{t_0})}. \tag{2.13}$$

as well as

$$X_t = \mathrm{e}^{-\left(\kappa+\frac{\alpha^2}{2}\right)t+\alpha W_t} \cdot \left[X_0 + \int_{t_0}^{t} \kappa\theta \mathrm{e}^{\left(\kappa+\frac{\alpha^2}{2}\right)s-\alpha W_s} \mathrm{d}s \right] . \tag{2.14}$$

The functional form of solution (2.14) is somewhat reminiscent of the payoff function of a continuously monitored Asian option in a standard Black-Scholes framework, and thus it may be possible to derive the Laplace transform of the distribution of X_t analytically following the lead given by Geman and Yor [GY93]. However, whilst this is noteworthy in its own right, it is unlikely to aid in the development of fast and efficient numerical integration schemes for Monte Carlo simulations, especially if the ultimate aim is to use the process X to drive the diffusion coefficient in a second stochastic differential equation.

Beyond the cases $q = 0$, $q = 1/2$, and $q = 1$, as far as we know, there are no analytical or semi-analytical solutions. Nevertheless, we are able to discuss the boundary behaviour solely based on our knowledge of the drift and diffusion terms:

1. 0 is an attainable boundary for $0 < q < 1/2$ and for $q = 1/2$ if $\kappa\theta < \alpha^2/2$
2. 0 is unattainable for $q > 1/2$
3. ∞ is unattainable for all $q > 0$.

These statements can be confirmed by the aid of Feller's boundary classification which can be found in [KT81]. The stationary distribution of this process can be calculated as (see [AP04, Prop. 2.4])

$$\pi(y) = C(q)^{-1} y^{-2q} e^{M(y,q)}, \qquad C(q) = \int_0^\infty y^{-2q} e^{M(y,q)} \mathrm{d}y \tag{2.15}$$

with the auxiliary function $M(y, q)$ given by

1. $q = 1/2$
$$M(y,q) = \frac{2\kappa}{\alpha^2}\left(\theta \ln(y) - y\right) \tag{2.16}$$
2. $q = 1$
$$M(y,q) = \frac{2\kappa}{\alpha^2}\left(-\theta/y - \ln(y)\right) \tag{2.17}$$
3. $0 < q < 1/2$ and $1/2 < q < 1$
$$M(y,q) = \frac{2\kappa}{\alpha^2}\left(\frac{\theta y^{1-2q}}{1-2q} - \frac{y^{2-2q}}{2-2q}\right). \tag{2.18}$$

The above equations can be derived from the Fokker-Planck equation which leads to an ordinary differential equation of Bernoulli type. The first moment of the process (2.2) is given by

$$\mathsf{E}[V_t] = (V_{t_0} - \theta)e^{-\kappa t} + \theta \, . \tag{2.19}$$

We can also calculate the second moment for $q = 1/2$ or $q = 1$:

$$\mathsf{E}\left[V_t^2\right] = \begin{cases} \frac{\mathrm{e}^{-2\kappa t}\left(\mathrm{e}^{\kappa t}-1\right)\left(2V_0+\left(\mathrm{e}^{\kappa t}-1\right)\theta\right)\left(\alpha^2+2\theta\kappa\right)}{2\kappa} & \text{for} \quad q = 1/2 \\ \frac{2\mathrm{e}^{-2\kappa t}\theta\kappa\left(\mathrm{e}^{2\kappa t}\theta\left(\kappa-\alpha^2\right)+\mathrm{e}^{\kappa t}(V_0-\theta)\left(2\kappa-\alpha^2\right)+\mathrm{e}^{\alpha^2 t}\left(V_0\left(\alpha^2-2\kappa\right)+\theta\kappa\right)\right)}{\alpha^4-3\alpha^2\kappa+2\kappa^2} & \text{for} \quad q = 1 \, . \end{cases} \tag{2.20}$$

This means that in the case $q = 1$, for $\alpha^2 > \kappa$ which is typically required in order to calibrate to the market observable strongly pronounced implied-Black-volatility skew, the variance of volatility grows unbounded, despite the fact that the model appears to be mean-reverting. For long dated options, this is a rather undesirable feature to have. On the other hand, in the case $q = 1/2$, for $\alpha^2 > \kappa$, instantaneous variance can attain zero, which is also undesirable for economical reasons. In addition to that, for the modelling of path dependent derivatives, the model (2.2) requires the use of numerical integration schemes that preserve the analytical properties of the variance process such as to remain on the real axis, or to simply stay positive. In the next section, we discuss alternatives for the generation of the stochastic volatility process that make the integration of volatility itself practically trivial.

2.2 Transformed Ornstein-Uhlenbeck

The origin of this process goes back to Uhlenbeck and Ornstein's publication [UO30] in which they describe the velocity of a particle that moves in an environment with friction. Doob [Doo42] first treated this process purely mathematically and expressed it in terms of a stochastic differential equation. In modern financial mathematics, the use of Ornstein-Uhlenbeck processes is almost commonplace. The attractive features of an Ornstein-Uhlenbeck process are that, whilst it provides a certain degree of flexibility over its auto-correlation structure, it still allows for the full analytical treatment of a standard Gaussian process.

In this article, we chose the formulation (2.4) to describe the Ornstein-Uhlenbeck process since we prefer a parametrisation that permits complete separation between the mean reversion speed and the variance of the *limiting* or *stationary distribution* of the process. The solution of (2.4) is

$$Y_t = e^{-\kappa t}\left(y_0 + \int_0^t e^{\kappa u}\alpha\sqrt{2\kappa}\,\mathrm{d}Z_u\right) \tag{2.21}$$

with initial time $t_0 = 0$. In other words, the stochastic process at time t is Gaussian with

$$Y_t \sim \mathcal{N}\left(y_0 \mathrm{e}^{-\kappa t}, \alpha^2\left(1 - \mathrm{e}^{-2\kappa t}\right)\right) \tag{2.22}$$

and thus the stationary distribution is Gaussian with variance α^2: a change in parameter κ requires no rescaling of α if we wish to hold the long-term uncertainty in the process unchanged. It is straightforward to extend the above results to the case when $\kappa(t)$ and $\alpha(t)$ are functions of time [Jäc02]. Since the variance of the driving Ornstein-Uhlenbeck process is the main criterion that determines the uncertainty in volatility for the financial underlying process, all further considerations are primarily expressed in terms of

$$\eta(t) \;:=\; \alpha \cdot \sqrt{1 - \mathrm{e}^{-2\kappa t}}\,. \tag{2.23}$$

There are fundamental differences between the requirements in the financial modelling of underlying asset prices, and the modelling of instantaneous stochastic volatility, or indeed any other not directly market-observable quantity. For reasons of financial consistency, we frequently have to abide by no-arbitrage rules that impose a specific functional form for the instantaneous drift of the underlying. In contrast, the modelling of stochastic volatility is typically more governed by long-term realism and structural similarity to real-world dynamics, and no externally given drift conditions apply. No-arbitrage arguments and their implied instantaneous drift conditions are omnipresent in financial arguments, and as a consequence, most practitioners have become used to thinking of stochastic processes exclusively in terms of an explicit stochastic differential equation. However, when there are no explicitly given conditions on the instantaneous drift, it is, in fact, preferable to model a stochastic process in the most analytically convenient form available. In other words, when preferences as to the attainable domain of the process are to be considered, it is in practice much more intuitive to start with a simple process of full analytical tractability, and to *transform* its domain to the target domain by virtue of a simple analytical function. For the modelling of stochastic volatility, this means that we utilise the flexible yet

tractable nature of the Ornstein-Uhlenbeck process (2.4) in combination with a strictly monotonic and differentiable mapping function $f : \mathbb{R} \to \mathbb{R}_+$.

One simple analytical transformation we consider is the exponential function, and the resulting stochastic volatility model was first proposed in [Sco87, equation (7)]. The model is intuitively very appealing: for any future point in time, volatility has a lognormal distribution which is a very comfortable distribution for practitioners in financial mathematics. Alas, though, recent research [AP04] has cast a shadow on this model's analytical features. It appears that, in its full continuous formulation, the log-normal volatility model can give rise to unlimited higher moments of the underlying financial asset. However, as has been discussed and demonstrated at great length for the very similar phenomenon of infinite futures returns when short rates are driven by a lognormal process [HW93, SS94, SS97a, SS97b], this problem vanishes as soon as the continuous process model is replaced by its discretised approximation which is why lognormal volatility models remain numerically tractable in applications. Still, in order to avoid this problem altogether, we introduce an alternative to the exponential transformation function which is given by a simple hyperbolic form.

In the following, we refer to

$$f_{\text{exp}}(y) \;:=\; \mathrm{e}^{y} \qquad\qquad \sigma_{\text{exp}}(y) \;:=\; \sigma_0 \cdot f_{\text{exp}}(y) \tag{2.24}$$

as the *exponential volatility* transformation also known as Scott's model [Sco87], and to

$$f_{\text{hyp}}(y) \;:=\; y + \sqrt{y^2 + 1} \qquad\qquad \sigma_{\text{hyp}}(y) \;:=\; \sigma_0 \cdot f_{\text{hyp}}(y) \tag{2.25}$$

as the *hyperbolic volatility* transformation. The densities of the exponential and hyperbolic volatilities are given by

$$\psi_{\text{exp}}(\sigma_{\text{exp}}, \sigma_0, \eta) \;=\; \frac{\varphi\left(f_{\text{exp}}^{-1}(\sigma_{\text{exp}}/\sigma_0), \eta\right)}{\mathrm{d}\sigma_{\text{exp}}/\,\mathrm{d}y} \tag{2.26}$$

$$\psi_{\text{hyp}}(\sigma_{\text{hyp}}, \sigma_0, \eta) \;=\; \frac{\varphi\left(f_{\text{hyp}}^{-1}(\sigma_{\text{hyp}}/\sigma_0), \eta\right)}{\mathrm{d}\sigma_{\text{hyp}}/\,\mathrm{d}y} \tag{2.27}$$

with

$$f_{\text{exp}}^{-1}(\sigma/\sigma_0) \;=\; \ln(\sigma/\sigma_0) \tag{2.28}$$

$$f_{\text{hyp}}^{-1}(\sigma/\sigma_0) \;=\; (\sigma/\sigma_0 - \sigma_0/\sigma)/\,2 \tag{2.29}$$

$$\mathrm{d}\sigma_{\text{exp}}/\,\mathrm{d}y \;=\; \sigma_{\text{exp}} \tag{2.30}$$

$$\mathrm{d}\sigma_{\text{hyp}}/\,\mathrm{d}y \;=\; \frac{2\sigma_0\sigma_{\text{hyp}}^2}{\sigma_0^2 + \sigma_{\text{hyp}}^2} \tag{2.31}$$

and

$$\varphi(y, \eta) \;:=\; \frac{\mathrm{e}^{-\frac{1}{2}\left(\frac{y}{\eta}\right)^2}}{\eta \cdot \sqrt{2\pi}}\,. \tag{2.32}$$

The hyperbolic transformation has been chosen to match the exponential form as closely as possible near the origin, and only to differ significantly in the regions of lower probability given for $|y/\eta| > 1$. The functional forms of the exponential and hyperbolic transformation are shown in comparison in figure 1.

2.2.1 Exponential vs. Hyperbolic transformation

In figure 2, we compare the densities of the Ornstein-Uhlenbeck process transformed with (2.24) and (2.25) given by equations (2.26) and (2.27). At first glance on a linear scale, we see a reasonable similarity between the two distributions. However, on a logarithmic scale, the differences in the tails of the distributions become clear: the hyperbolic transformation has significantly lower probability for both very low values as well as for large values.

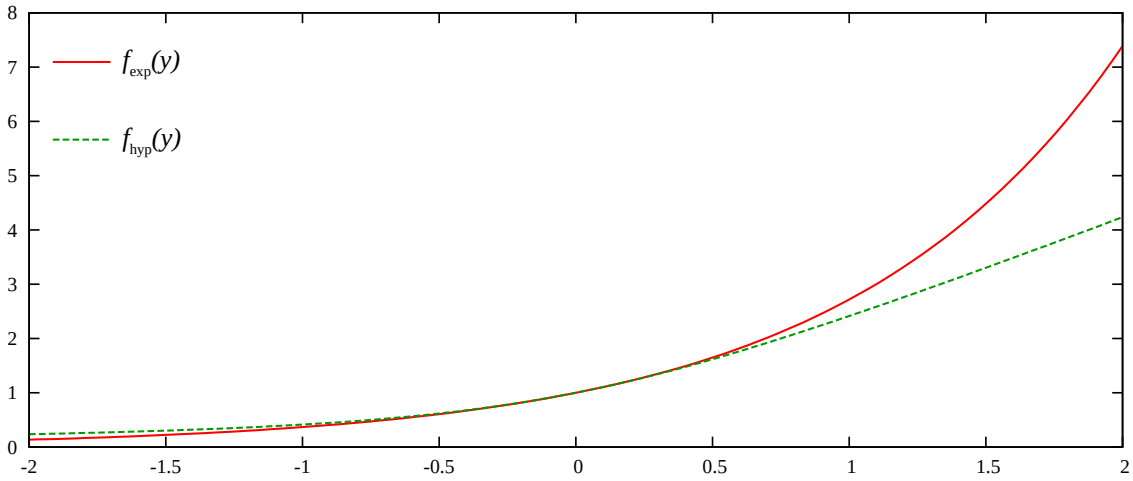

Figure 1: The exponential and hyperbolic transformation functions.

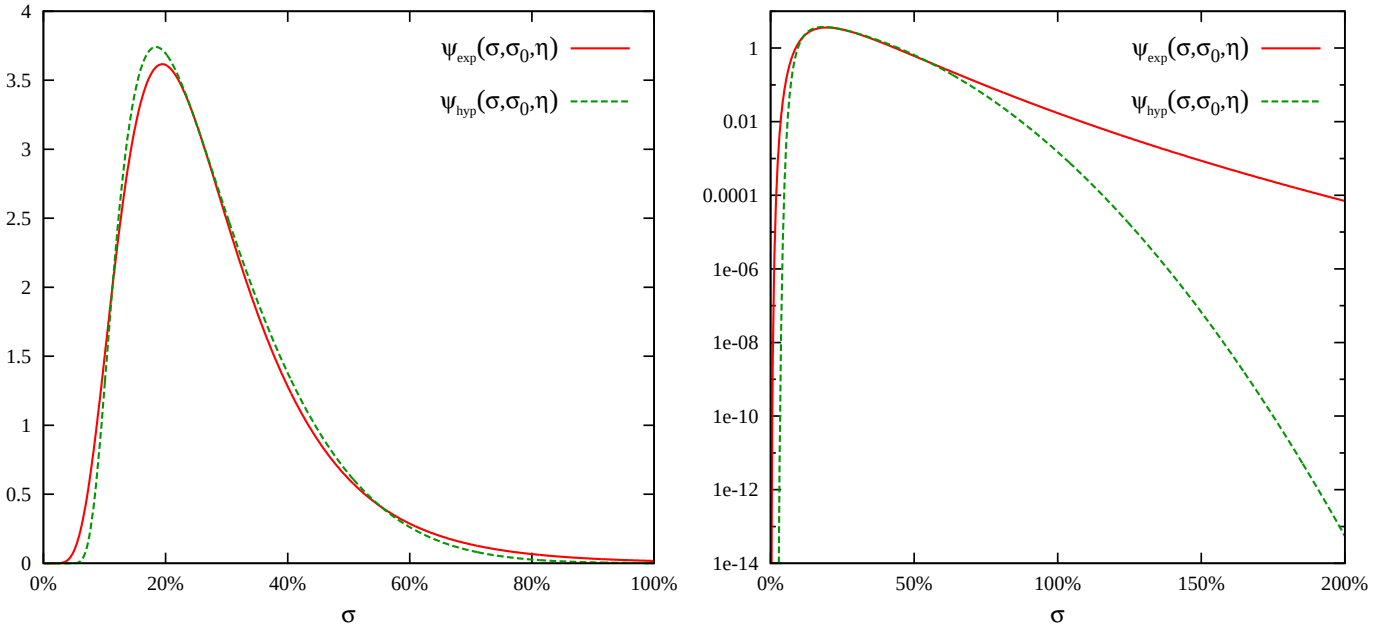

Figure 2: Densities of instantaneous volatility using the exponential and the hyperbolic transformation of the driving Ornstein-Uhlenbeck process for $\sigma_0 = 25\%$ and $\eta = 1/2$. Note the distinctly different tails of the distributions.

Returning to the analytical form of the density functions (2.26) and (2.27), it is interesting to note that, given that $y = f^{-1}(\sigma/\sigma_0)$ is Gaussian, the volatility distribution implied by the inverse hyperbolic transformation (2.29) is nearly Gaussian for large values of instantaneous volatility $\sigma \gg 0$ since we have

$$\sigma_{hyp} \approx 2\sigma_0 \cdot y \quad \text{for} \quad \sigma_{hyp} \gg 0\,. \tag{2.33}$$

This feature is particularly desirable since it ensures that the tail of the volatility distribution at the higher end is as thin as the Gaussian process itself, and thus no moment explosions are to be feared for the underlying. Conversely, for small values of instantaneous volatility $\sigma \ll 1$, the volatility distribution implied by the hyperbolic volatility model is nearly *inverse* Gaussian because of

$$\sigma_{hyp} \approx -\sigma_0/2/\,y \quad \text{for} \quad \sigma_{hyp} \to 0\,. \tag{2.34}$$

In a certain sense, the hyperbolic model can be seen as a blend of an inverse Gaussian model at the lower end of the distribution, and a Gaussian density at the upper end, with their respectively thin tails. In contrast, exponential volatility is simply lognormally distributed, which in turn gives rise to distinctly fatter tails than the normal (at the high end) or inverse normal (at the low end) density.

From the basis of our complete analytical understanding of both the exponential and the hyperbolic volatility process, we can use Itô's lemma to derive their respective stochastic differential equations. For

the exponential transformation (2.24) we obtain Scott's original SDE [Sco87, equation (7)]

$$\mathrm{d}\sigma \;=\; \kappa\sigma\left(\alpha^2 - \ln(\sigma/\sigma_0)\right)\mathrm{d}t + \alpha\sigma\sqrt{2\kappa}\,\mathrm{d}Z\;. \tag{2.35}$$

It is remarkable to see the difference in the stochastic differential equation we obtain for the hyperbolic volatility process (2.25):

$$\mathrm{d}\sigma \;=\; -\kappa\sigma\frac{\sigma^6+\sigma^4\sigma_0^2-(8\alpha^2+1)\sigma^2\sigma_0^4-\sigma_0^6}{\left(\sigma^2+\sigma_0^2\right)^3}\,\mathrm{d}t \;+\; \sigma\alpha\sqrt{8\kappa}\frac{\sigma\sigma_0}{\sigma^2+\sigma_0^2}\,\mathrm{d}Z\;. \tag{2.36}$$

The complexity of the explicit form (2.36) of the hyperbolic volatility process may help to explain why it has, to the best of our knowledge, not been considered before. As we know, though, the resulting process is of remarkable simplicity and very easy to simulate directly, whilst overcoming some of the weaknesses of the (calibrated) CIR/Heston process (namely that zero is attainable), as well as the moment divergence when volatility or variance is driven by lognormal volatility as incurred by the Brennan-Schwartz process for volatility and Scott's model.

The moments of the exponential transformation function are

$$\mathsf{E}[(f_{\text{exp}}(y))^m] \;=\; \mathrm{e}^{\frac{1}{2}m^2\eta^2}\;. \tag{2.37}$$

For the hyperbolic transformation we obtain the general solution

$$\mathsf{E}[(f_{\text{hyp}}(y))^m] \;=\; \frac{2^{\frac{3}{2}n}}{\sqrt{4\pi}}\eta^n\Gamma\left(\tfrac{1+n}{2}\right){}_1F_1(-\tfrac{n}{2},1-n,\tfrac{1}{2\eta^2}) \;+\; \frac{2^{-\frac{3}{2}n}}{\sqrt{4\pi}}\eta^{-n}\Gamma\left(\tfrac{1-n}{2}\right){}_1F_1(\tfrac{n}{2},1+n,\tfrac{1}{2\eta^2}) \tag{2.38}$$

in terms of Kummer's hypergeometric function ${}_1F_1$. The first two moments, specifically, are given by

$$\mathsf{E}\left[(f_{\text{hyp}}(y))^1\right] \;=\; \sqrt{2}\cdot\eta\cdot U(-\tfrac{1}{2},0,\tfrac{1}{2\eta^2}) \tag{2.39}$$

$$\mathsf{E}\left[(f_{\text{hyp}}(y))^2\right] \;=\; 1+2\eta^2 \tag{2.40}$$

where $U(a,b,z)$ is the logarithmic confluent hypergeometric function. More revealing than the closed form for the moments of the respective transformation functions is an analysis based on their Taylor expansions

$$f_{\text{exp}}(y) \;=\; 1+y+\frac{y^2}{2}+\mathcal{O}\left(y^3\right) \tag{2.41}$$

$$f_{\text{hyp}}(y) \;=\; 1+y+\frac{y^2}{2}+\mathcal{O}\left(y^4\right)\;. \tag{2.42}$$

Thus, for both of these functions,

$$(f_{\dots}(y))^n \;=\; 1+n\cdot y+\left[\binom{n}{2}+\frac{n}{2}\right]\cdot y^2+\mathcal{O}\left(y^3\right)\;. \tag{2.43}$$

Since y is normally distributed with mean 0 and variance η^2, and since all odd moments of the Gaussian distribution vanish, this means that for both the exponential and the hyperbolic transformation we have

$$\mathsf{E}[(f_{\dots}(y))^n] \;=\; 1+\left[\binom{n}{2}+\frac{n}{2}\right]\cdot\eta^2+\mathcal{O}\left(\eta^4\right)\;. \tag{2.44}$$

The implication of (2.44) is that *all* moments of the exponential and the hyperbolic transformation function agree up to order $\mathcal{O}\left(\eta^3\right)$. We show an example for this in figure 3. As η increases, the moments of the exponential function grow faster by a term of order $\mathcal{O}\left(\eta^4\right)$.

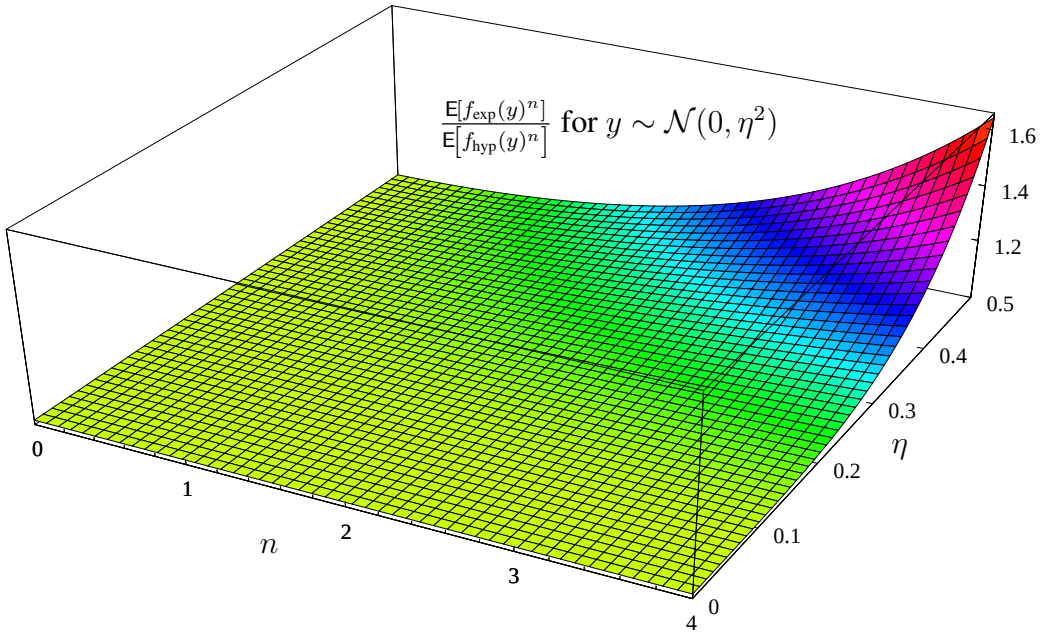

Figure 3: Comparison of the moments of the exponential and the hyperbolic transformation functions. Note that the floor level is exactly 1.

3 Numerical integration of mean-reverting CEV processes

The numerical integration of the coupled stochastic volatility system (2.1) and (2.2) is composed of two different parts. First, we have to find an appropriate method for the approximation of the stochastic volatility process itself, and secondly we need to handle the dynamics of the financial underlying (2.1) whose diffusion part is affected by the stochasticity of volatility.

Since the volatility process does not explicitly depend on the underlying, we can treat it separately. In order to retain numerical stability and to achieve good convergence properties, it is desirable for the numerical integration scheme of the volatility or variance process to preserve structural features such as positivity. For the exponentially or hyperbolically transformed Ornstein-Uhlenbeck process, this is trivially taken care of by the transformation function itself. For the mean-reverting CEV process (2.2), however the design of a positivity preserving scheme is a task in its own right. The simplest approach, for instance, namely the explicit Euler scheme

$$X_{n+1} = X_n + \kappa(\theta - X_n)\Delta t_n + \alpha X_n^q \Delta W_n \ , \tag{3.1}$$

fails to preserve positivity. The same deficiency is exhibited by the standard Milstein and the Milstein+ scheme whose formulæ we give in appendices A.1 and A.2, respectively. The Balanced Implicit Method (BIM) as introduced by Milstein, Platen and Schurz [MPS98], however,

$$\begin{aligned} X_{n+1} &= X_n + \kappa(\theta - X_n)\Delta t_n + \alpha X_n^q \Delta W_n + C(X_n)(X_n - X_{n+1}) && (3.2)\\ C(X_n) &= c_0^{\mathrm{BIM}}(X_n)\Delta t_n + c_1^{\mathrm{BIM}}(X_n)|\Delta W_n| && (3.3) \end{aligned}$$

with control functions

$$\begin{aligned} c_0^{\mathrm{BIM}}(x) &= \kappa && (3.4)\\ c_1^{\mathrm{BIM}}(x) &= \alpha x^{1-q} && (3.5) \end{aligned}$$

is able to preserve positivity as is shown in [Sch96]. Alas, this scheme only achieves the same strong order of convergence as the Euler scheme, i.e. $1/2$. This means, whilst the Balanced Implicit Method helps to overcome the problem of spurious negative values for variance, it does not increase the convergence speed. In fact, when a step size is chosen such that for the specific set of parameters at hand the explicit

Euler scheme is usable[1], the Balanced Implicit Method often has worse convergence properties than the explicit Euler method. This feature of the Balanced Implicit Method is typically caused by the fact that the use of the weight function c_1^{BIM} effectively increases the unknown coefficient dominating the leading error terms.

Another scheme that has been shown to preserve positivity for certain parameter ranges is the adaptive Milstein scheme [Kah04] with suitable stepsize $\Delta\tau_n$ and $\tilde{z} \sim \mathcal{N}(0,1)$

$$X_{n+1} = X_n + \kappa(\theta - X_n)\Delta\tau_n + \alpha X_n^q \sqrt{\Delta\tau_n}\tilde{z} + \frac{1}{2}\alpha^2 q X_n^{2q-1}\Delta\tau_n\left(\tilde{z}^2 - 1\right) . \tag{3.6}$$

Unfortunately, this scheme requires adaptive resampling and thus necessitates the use of a pseudo-random number pipeline which in turn disables or hinders a whole host of independently available convergence enhancement techniques such as low-discrepancy numbers, importance sampling, stratification, latin-hypercube methods, etc. An advanced method that obviates the use of pseudo-random number pipelines is based on the combination of the Milstein scheme with the idea of balancing: the Balanced Milstein Method (BMM)

$$\begin{aligned} X_{n+1} &= X_n + \kappa(\theta - X_n)\Delta t_n + \alpha X_n^q \Delta W_n + \frac{1}{2}\alpha^2 q X_n^{2q-1}\left(\Delta W_n^2 - \Delta t_n\right) \\ &\quad + D(X_n)\left(X_n - X_{n+1}\right), \end{aligned} \tag{3.7}$$

$$D(X_n) = d_0^{\text{BMM}}(X_n)\Delta t_n + d_1^{\text{BMM}}(X_n)\left(\Delta W_n^2 - \Delta t_n\right). \tag{3.8}$$

As in the Balanced Implicit Method we can control the integration steps by using weighting functions $d_0^{\text{BMM}}(\cdot)$ and $d_1^{\text{BMM}}(\cdot)$. The choice of these weighting functions strongly depends on the structure of the SDE. It can be shown (see [KS05, Theorem 5.9]) that the BMM preserves positivity for the mean-reverting CEV model (2.2) with the following choice

$$d_0^{\text{BMM}}(x) = \Theta\kappa + \frac{1}{2}\alpha^2 q |x|^{2q-2} , \tag{3.9}$$

$$d_1^{\text{BMM}}(x) = 0 . \tag{3.10}$$

The parameter $\Theta \in [0,1]$ provides some freedom for improved convergence speed but it has to be chosen such that

$$\Delta t_n < \frac{2q-1}{2q\kappa(1-\Theta)} . \tag{3.11}$$

It is always safe to choose $\Theta = 1$, though, for improved performance, we used $\Theta = 1/2$ whenever this choice was possible[2].

The above mentioned integration methods, namely the standard explicit Euler scheme, the Balanced Implicit Method, the Balanced Milstein Method, and the adaptive Milstein scheme, deal with the stochastic differential equation in its original form (2.2). Another approach to integrate (2.2) whilst preserving positivity is to transform the stochastic differential equation to logarithmic coordinates using Itô's lemma as suggested by Andersen and Brotherton-Ratcliffe [ABR01]. Applying this to the mean-reverting CEV process leads to

$$\mathrm{d}\ln V_t = \frac{2\kappa(\theta - V_t) - \alpha^2 V_t^{2q-1}}{2V_t}\mathrm{d}t + \alpha V_t^{q-1}\mathrm{d}Z_t \tag{3.12}$$

which can be solved by the aid of a simple Euler scheme. The major drawback with this approach is that, whilst the Euler scheme applied to the transformed stochastic differential equation (3.12) preserves

[1] For most schemes, spurious negative values incurred as an undesirable side effect of the numerical method disappear as the step size Δt is decreased. For a negative variance to appear at any one step, the drawn normal variate generating the step typically has to exceed a certain threshold. This threshold tends to grow as step size decreases. Thus, with decreasing step size, eventually, the threshold exceeds the maximum standard normal random number attainable on the finite representation computer system used.

[2] For $q = 1/2$ the numerator becomes zero. Despite this, positivity can be preserved with $d_0^{\text{BMM}} = \kappa$.

positivity, it is also likely to become unstable for suitable time steps [ABR01]. These instabilities are a direct consequence of the divergence of both the drift and the diffusion terms near zero. For that reason Andersen and Brotherton-Ratcliffe suggested a moment matched log-normal approximation

$$V_{n+1} = \left(\theta + (V_n - \theta)\,\mathrm{e}^{-\kappa\Delta t_n}\right)\mathrm{e}^{-\frac{1}{2}\Gamma_n^2 + \Gamma_n \tilde{z}}\,, \tag{3.13}$$

$$\Gamma_n = \ln\left(1 + \frac{\frac{1}{2}\alpha^2 V_n^{2p}\kappa^{-1}\left(1 - e^{-2\kappa\Delta t_n}\right)}{\left(\theta + (V_n - \theta)\,\mathrm{e}^{-\kappa\Delta t_n}\right)^2}\right) \tag{3.14}$$

with $\tilde{z} \sim \mathcal{N}(0,1)$. We will refer to this integration scheme as *moment matched Log-Euler* in the following. This method is at its most effective for the Brennan-Schwartz model (3.23) as we can see in figure 7 (B) since for $p = 1$ the logarithmic transformation leads to an additive diffusion (3.12) term. However, even in that case, it is outperformed by the bespoke method we call *Pathwise Adapted Linearisation* which is explained in section 3.1, as well as the Balanced Milstein method (3.7). For the Heston case, where the stochastic volatility is given by the Cox-Ingersoll-Ross equation with $q = 1/2$ which is shown in figures 5 and 6, the moment matched log-Euler method has practically no convergence advantage over straightforward explicit Euler integration as long as the size of α is reasonably small. Contrarily, the approximation quality of all integration schemes is decisively reduced when dealing with large α as we can see in figure 5 (B). Making matters even worse, one can observe that schemes of Milstein type are losing their strong convergence order of 1. The explanation for this behaviour is rather simple: the Milstein method is not even guaranteed to converge at all for the mean-reverting CEV process (2.2)! Having a closer look at the diffusion $b(x) = \alpha x^q$, we recognize that for $q < 1$ this function is not continuously differentiable on $\mathbb{R}$ which is necessary for the application of stochastic Taylor expansion techniques. Nonetheless, as long as the stochastic process is analytically positive, i.e. $x > 0$ there exists a local stochastic Taylor expansion preserving strong convergence of the Milstein method. However, when zero is attainable, the discontinuity of the first derivative of the diffusion $b(x)$ reduces the strong convergence order to $1/2$.

In figures 5, 6, and 7 we present examples for the convergence behaviour of the different methods in comparison. For the standard Milstein (A.5) and the Milstein+ scheme (A.15), for some of the parameter configurations, it was necessary to floor the simulated variance values at zero since those schemes do not preserve positivity by construction.

The depicted strong approximation convergence measure is given by the L_2 norm of the difference between the simulated terminal value, and the terminal value of the reference calculation, averaged over all M paths, i.e.

$$\sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(X_i^{(n_{\text{steps}})}(T) - X_i^{(n_{\text{reference}})}(T)\right)^2}\,. \tag{3.15}$$

This quantity is shown as a function of average CPU time per path. This was done because the ultimate criterion for the choice of any integration method in applications is the cost of accuracy in terms of calculation time since calculation time directly translates into the amount of required hardware for large scale computations such as overnight risk reports, or into user downtime when interactive valuations are needed. This does, of course, make the results dependent on the used hardware[3], not only in absolute terms but also in relative terms since different processor models require different numbers of CPU clock cycles for all the involved basic floating point operations. Nevertheless, the pathwise error as a function of average CPU time is probably the most significant criterion for the quality of any integration method. Examples for this consideration are the fact that in figure 6 the nominal advantage of the *moment matched Log-Euler* is almost precisely offset by the additional calculation time it requires compared to the Euler scheme, and also that in figure 7 (B) the relative performance of the Balanced Milstein Method is compatible with the scheme denoted as *Pathwise Adapted Linearisation* which is explained in section 3.1.1.

[3]Throughout this article, all calculations shown were carried out on a processor from the Intel Pentium series (Family 6, Model 9, Stepping 5, Brand id 6, CPU frequency 1700 MHz).

The curves in figures 5, 6 and 7 have been constructed by repeated simulation with increased numbers of steps in the Brownian bridge Wiener path generation in powers of two from 1 to 128:

$$n_{\text{steps}} \in \{1, 2, 4, 8, 16, 32, 64, 128\} \; . \tag{3.16}$$

The reference solution was always computed with 2^{15} steps. The number generation mechanism used was the Sobol' algorithm [Jäc02] throughout apart from figure 6 (B) where we also show the results from using the Mersenne Twister [MN98] in comparison. Note that the results are fairly insensitive to the choice of number generator. In addition to the methods discussed above, we also included the results from bespoke schemes denoted as *Pathwise Adapted Linearisation*. These schemes are carefully adapted to the respective equation and we introduce them in the following section.

3.1 Pathwise approximations for specific cases

Yet another approach for the numerical integration of stochastic differential equations of the form

$$\mathrm{d}X \;=\; a(X)\mathrm{d}t + b(X)\mathrm{d}Z \tag{3.17}$$

as it is the case for (2.2) is to apply Doss's [Dos77] method of constructing pathwise solutions first used in the context of numerical integration schemes by Pardoux and Talay [PT85]. The formal derivation of Doss's pathwise solution can be found in [KS91, pages 295–296].

In practice, Doss's method can hardly ever be applied directly since it is essentially just an existence theorem that states that any process for which there is a unique strong solution can be seen as a transformation of the solution to an ordinary differential equation with a stochastic inhomogeneity, i.e. a solution of the form

$$X \;=\; f(Y, Z) \quad \text{with boundary condition} \quad f(Y, Z_0) = Y \tag{3.18}$$

with

$$\mathrm{d}Y \;=\; g(Y, Z)\mathrm{d}t \tag{3.19}$$

implying

$$X_0 \;=\; Y_0 \tag{3.20}$$

whereby the functions f and g can be derived constructively from the stochastic differential equation for X:

$$\partial_Y f(Y, Z) \;=\; \mathrm{e}^{\int_{Z_0}^{Z} b'(f(y,z))\,\mathrm{d}z} \tag{3.21}$$

$$g(Y, Z) \;=\; \left[a\left(f(Y, Z)\right) - \frac{1}{2} \cdot b\left(f(Y, Z)\right) \cdot b'\left(f(Y, Z)\right)\right] \cdot \mathrm{e}^{-\int_{Z_0}^{Z} b'(f(Y,z))\,\mathrm{d}z} \; . \tag{3.22}$$

Even though one can rarely use Doss's method in its full analyticity, one can often devise a powerful bespoke approximate discretisation scheme for the stochastic differential equation at hand based on Doss's pathwise existence theorem by the aid of some simple approximative assumptions without the need to go through the Doss formalism itself.

3.1.1 Pathwise approximation of the Brennan-Schwartz SDE

For $q = 1$, the mean-reverting CEV process (2.2) becomes

$$\mathrm{d}X = \kappa(\theta - X)\mathrm{d}t + \alpha X \mathrm{d}Z \; . \tag{3.23}$$

Assuming

$$\kappa > 0 \; , \quad \theta > 0 \; , \quad \alpha > 0 \; , \quad \text{and} \quad X(0) > 0 \; , \tag{3.24}$$

we must have

$$X_t \geq 0 \qquad \text{for all} \qquad t > 0 \; . \tag{3.25}$$

Using equation (3.21), we obtain

$$f(Y, Z) = Y \mathrm{e}^{\alpha Z} \; . \tag{3.26}$$

and by the aid of (3.22), we have

$$\mathrm{d}Y \;=\; \left[\kappa\theta \mathrm{e}^{-\alpha Z} - \left(\kappa + \frac{1}{2}\alpha^2\right) Y\right] \mathrm{d}t \; . \tag{3.27}$$

We cannot solve this equation directly. Also, a directly applied explicit Euler scheme would permit Y to cross over to the negative half of the real axis and thus $X = f(Y, Z) = Y\mathrm{e}^{\alpha Z}$ would leave the domain of (3.23). What's more, an explicit Euler scheme applied to equation (3.27) would mean that, within the scheme, we interpret Z_t as a piecewise constant function. Not surprisingly, it turns out below that we can do better than that!

Recall that, for the given time discretisation, we explicitly construct the Wiener process values $Z(t_i)$ and thus, for the purpose of numerical integration of equation (3.23), they are known along any one given path. If we now approximate Z_t as a piecewise linear function in between the known values at t_n and t_{n+1}, i.e.

$$Z_t \;\approx\; \beta_n t + \gamma_n \quad \text{for} \quad t \in [t_n, t_{n+1}] \tag{3.28}$$

with

$$\gamma_n \;=\; Z(t_n) - \beta_n t_n \qquad \text{and} \qquad \beta_n \;=\; \frac{Z(t_{n+1}) - Z(t_n)}{t_{n+1} - t_n} \; , \tag{3.29}$$

then we have the approximating ordinary differential equation

$$\mathrm{d}\hat{Y} \;=\; \left[\kappa\theta \mathrm{e}^{-\alpha(\beta_n t + \gamma_n)} - \left(\kappa + \tfrac{1}{2}\alpha^2\right) \hat{Y}\right] \mathrm{d}t \; . \tag{3.30}$$

Using the abbreviations

$$\delta_n := \kappa + \tfrac{1}{2}\alpha^2 - \alpha\beta_n \; , \qquad \Delta t_n := t_{n+1} - t_n \; , \qquad \text{and} \qquad Z_{n+1} := Z(t_{n+1})$$

we can write the solution to equation (3.30) as

$$\hat{Y}_{n+1} = \hat{Y}_n \mathrm{e}^{-\left(\kappa + \frac{1}{2}\alpha^2\right)\Delta t_n} + \kappa\theta \cdot \mathrm{e}^{-\alpha Z_{n+1}} \cdot \left(\frac{1 - \mathrm{e}^{-\delta_n \Delta t_n}}{\delta_n}\right) \; , \tag{3.31}$$

which gives us

$$\hat{X}_{n+1} = \hat{X}_n \mathrm{e}^{-\delta_n \Delta t_n} + \kappa\theta \cdot \left(\frac{1 - \mathrm{e}^{-\delta_n \Delta t_n}}{\delta_n}\right) \; . \tag{3.32}$$

This scheme is unconditionally stable. We refer to it as *Pathwise Adapted Linearisation* in the following. Apart from its stability, this scheme has the additional desirable property that, in the limits $\theta \to 0$ and/or $\kappa \to 0$, i.e. in the limit of equation (3.23) resembling a standard geometric Brownian motion, it is free of any approximation. Since in practice θ and/or κ tend to be not too large, the scheme's proximity to exactness translates into a remarkable acccuracy when used in applications.

It is interesting to note that a similar approach based on replacing the term $\mathrm{d}Z$ directly in the stochastic differential equation

$$\mathrm{d}X \;=\; \kappa(\theta - X)\mathrm{d}t + \alpha X \mathrm{d}Z \tag{3.33}$$

by a linear approximation $\mathrm{d}Z \approx \beta \mathrm{d}t$ gives rise to a scheme that *does not converge in the limit* $\Delta t \to 0$ as first observed by Wong and Zakai [WZ65]. However, if we make the same replacement in the Milstein scheme and drop terms of order $\mathcal{O}(\mathrm{d}t^2)$ and higher, which for (3.23) means

$$\Delta X \;\approx\; \kappa(\theta - X)\Delta t + \alpha X \Delta Z + \frac{1}{2}\alpha^2 X \left(\Delta Z^2 - \Delta t\right) \tag{3.34}$$

$$\Delta X \;\approx\; \kappa(\theta - X)\Delta t + \alpha X \beta \Delta t + \frac{1}{2}\alpha^2 X \left(\beta^2 \Delta t^2 - \Delta t\right) \tag{3.35}$$

$$\frac{\mathrm{d}X}{\mathrm{d}t} \;\approx\; \kappa(\theta - X) - \frac{1}{2}\alpha^2 X + \alpha\beta X \; , \tag{3.36}$$

and integrate, we arrive at exactly the same scheme (3.32) as if we had gone through the full Doss formalism. The reason for this is that the lowest order scheme that includes explicitly all terms that are individually in expectation of order $\mathrm{d}t$ is the Milstein scheme, not the Euler scheme, and the difference terms are crucial to preserve strong convergence when we introduce piecewise linearisation of the discretised Wiener process.

3.1.2 Pathwise approximation of the Cox-Ingersoll-Ross / Heston SDE

The special case $q = 1/2$ of (2.2) represents the stochastic differential equation of the variance process in the Heston model [Hes93], as well as the short rate process in the Cox-Ingersoll-Ross model [CIR85]

$$\mathrm{d}V = \kappa(\theta - V)\mathrm{d}t + \alpha\sqrt{V}\mathrm{d}Z \ . \tag{3.37}$$

In this case, an explicit solution of the Doss formalism (3.21) is not obvious. However, by conditioning on one specific path in Z we can bypass this difficulty by directly approximating Z_t as a piecewise linear function in between the known values as given in equations (3.28) and (3.29). Using the resulting dependency $\mathrm{d}Z = \beta_n \mathrm{d}t$ in the Milstein scheme applied to (3.37)

$$\mathrm{d}V \ \approx \ \kappa(\theta - V)\mathrm{d}t + \alpha\sqrt{V}\mathrm{d}Z + \frac{1}{4}\alpha^2\left(\mathrm{d}Z^2 - \mathrm{d}t\right) \ , \tag{3.38}$$

i.e.

$$\mathrm{d}V \ \approx \ \kappa(\theta - V)\mathrm{d}t + \alpha\sqrt{V}\beta_n\mathrm{d}t + \frac{1}{4}\alpha^2\left(\beta_n^2\mathrm{d}t^2 - \mathrm{d}t\right) \ , \tag{3.39}$$

and dropping terms of order $\mathrm{d}t^2$, we obtain the approximate ordinary differential equation

$$\frac{\mathrm{d}V}{\mathrm{d}t} \ \approx \ \kappa(\theta - V) - \frac{1}{4}\alpha^2 + \alpha\beta_n\sqrt{V} \tag{3.40}$$

which has the implicit solution

$$t - t_n \ = \ T(V_t, \beta_n) - T(V_{t_n}, \beta_n) \tag{3.41}$$

with

$$T(v,\beta) \ := \ \frac{2\alpha\beta}{\kappa\sqrt{\alpha^2\beta^2 + 4\theta\kappa^2 - \kappa\alpha^2}}\,\mathrm{atanh}\left(\frac{2\kappa\sqrt{v} - \alpha\beta}{\sqrt{\alpha^2\beta^2 + 4\theta\kappa^2 - \kappa\alpha^2}}\right) - \frac{1}{\kappa}\ln\left(\kappa\left(v - \theta\right) + \frac{1}{4}\alpha^2 - \alpha\beta\sqrt{v}\right) \ . \tag{3.42}$$

Equation (3.42) can be solved numerically comparatively readily since we know that, given β_n, over the time step from t_n to t_{n+1}, V_t will move monotonically, and that for all $\Delta t_n := (t_{n+1} - t_n)$ we have

$$V_{t_{n+1}} \ > \ \left(\frac{\alpha\left|\beta_n\right|}{2\kappa} - \sqrt{\left(\frac{\alpha\beta_n}{2\kappa}\right)^2 + \theta - \frac{\alpha^2}{4\kappa}}\right)^2 \tag{3.43}$$

which can be shown by setting the argument of the logarithm in the right hand side of equation (3.42) to zero. Alternatively, an inverse series expansion can be derived. Up to order $\mathcal{O}(\Delta t_n^4)$, we find

$$\begin{aligned} V_{n+1} \ = V_n + \left(\kappa(\tilde{\theta} - V_n) + \alpha\beta_n\sqrt{V_n}\right)\cdot\Delta t_n \cdot \\ \cdot\left[1 + \tfrac{\alpha\beta_n - 2\kappa\sqrt{V_n}}{4\sqrt{V_n}}\cdot\Delta t_n + \tfrac{\kappa\left(V_n\left(4\kappa\sqrt{V_n} - 3\alpha\beta_n\right) - \alpha\beta_n\tilde{\theta}\right)}{24\sqrt{V_n}^3}\cdot\Delta t_n^2\right. \\ \left. + \tfrac{\kappa\left(3\alpha\beta_n\kappa\tilde{\theta}^2 + \kappa V_n^2\left(7\alpha\beta_n - 8\kappa\sqrt{V_n}\right) + 2\alpha\beta_n\tilde{\theta}\sqrt{V_n}\left(\alpha\beta_n + \kappa\sqrt{V_n}\right)\right)}{192\sqrt{V_n}^5}\cdot\Delta t_n^3\right] + \mathcal{O}(\Delta t_n^5) \end{aligned} \tag{3.44}$$

with

$$\tilde{\theta} \;:=\; \theta - \frac{\alpha^2}{4\kappa}\;. \tag{3.45}$$

The shape of the curves generated by (3.42) and its 4th order inverse expansion (3.44) is shown in figure 4 where values for β directly represent the standard normal deviation equivalent of the drawn Gaussian random number. In the following, we denoted the expansion (3.44) as *Pathwise Adapted Linearisation Quartic*, and its second order truncation

$$V_{n+1} \;= V_n + \left(\kappa(\tilde{\theta} - V_n) + \alpha\beta_n\sqrt{V_n}\right)\cdot\Delta t_n \cdot \left[1 + \tfrac{\alpha\beta_n - 2\kappa\sqrt{V_n}}{4\sqrt{V_n}}\cdot\Delta t_n\right] + \mathcal{O}(\Delta t_n^3) \tag{3.46}$$

as *Pathwise Adapted Linearisation Quadratic*. We only show results for expansions of even order for reasons of numerical stability since all odd order expansion can reach zero which is undesirable. For small values of α as in figure 5 (A) both schemes are remarkable effective. Unfortunately, these schemes are inappropriate for large values of α due to numerical instabilities.

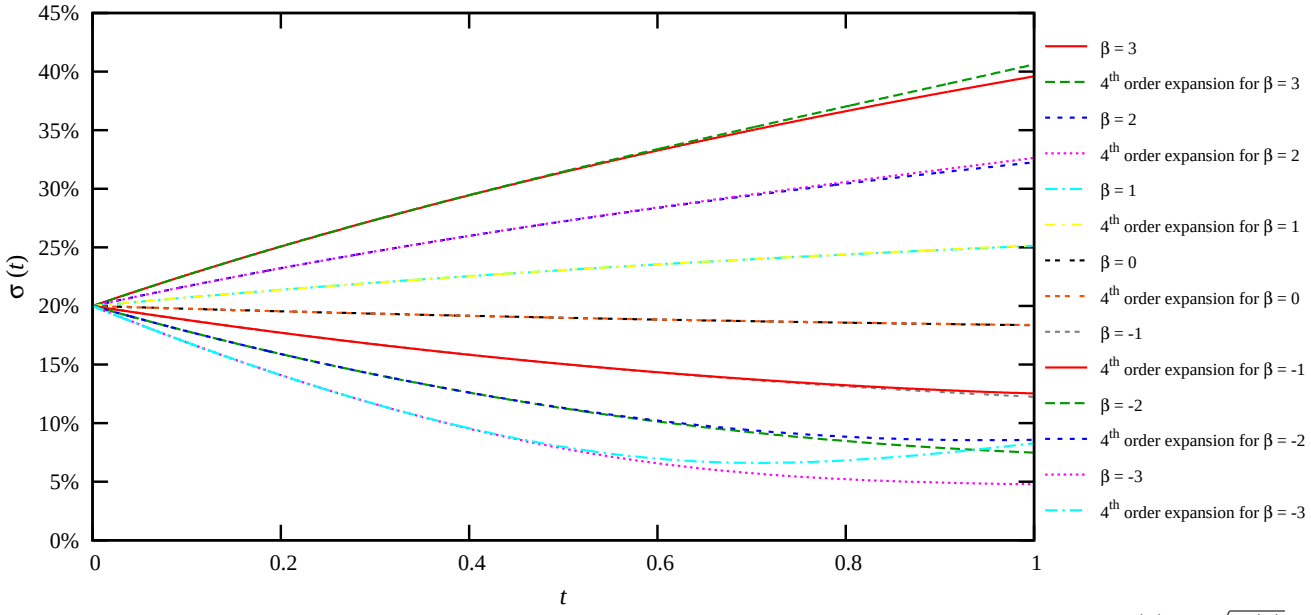

Figure 4: Approximation (3.42) and its quartic expansion (3.44) for the CIR/Heston volatility process for $\sigma(0) = \sqrt{V(0)} = 20\%$, $\theta = V(0)$, $\alpha = 20\%$, $\kappa = 1$ over a unit time step for different levels of the variate $\beta = Z(1) - Z(0)$.

4 Approximation of stochastic volatility models

In this section, we discuss the numerical treatment of the full two-dimensional stochastic volatility model. Irrespective of the volatility or variance process, the dynamics of the financial underlying are given by equation (2.1). As for the stochasticity of volatility/variance, both the transformed Ornstein-Uhlenbeck process as well as the mean-reverting CEV process (2.2) can be cast in the form

$$\mathrm{d}V_t = a(V_t)\mathrm{d}t + b(V_t)\mathrm{d}Z_t\;. \tag{4.1}$$

For the mean-reverting CEV process, the functional forms for a and b are directly given. For the exponentially and hyperbolically transformed Ornstein-Uhlenbeck process, they can be obtained from (2.35) and (2.36), respectively.

In logarithmic coordinates, the process equation for the financial underlying is given by

$$\ln S_t = \ln S_{t_0} + \int_{t_0}^{t}\mu(s)\mathrm{d}s - \tfrac{1}{2}\int_{t_0}^{t} V_s^{2p}\mathrm{d}s + \int_{t_0}^{t} V_s^{p}\mathrm{d}W_s\;. \tag{4.2}$$

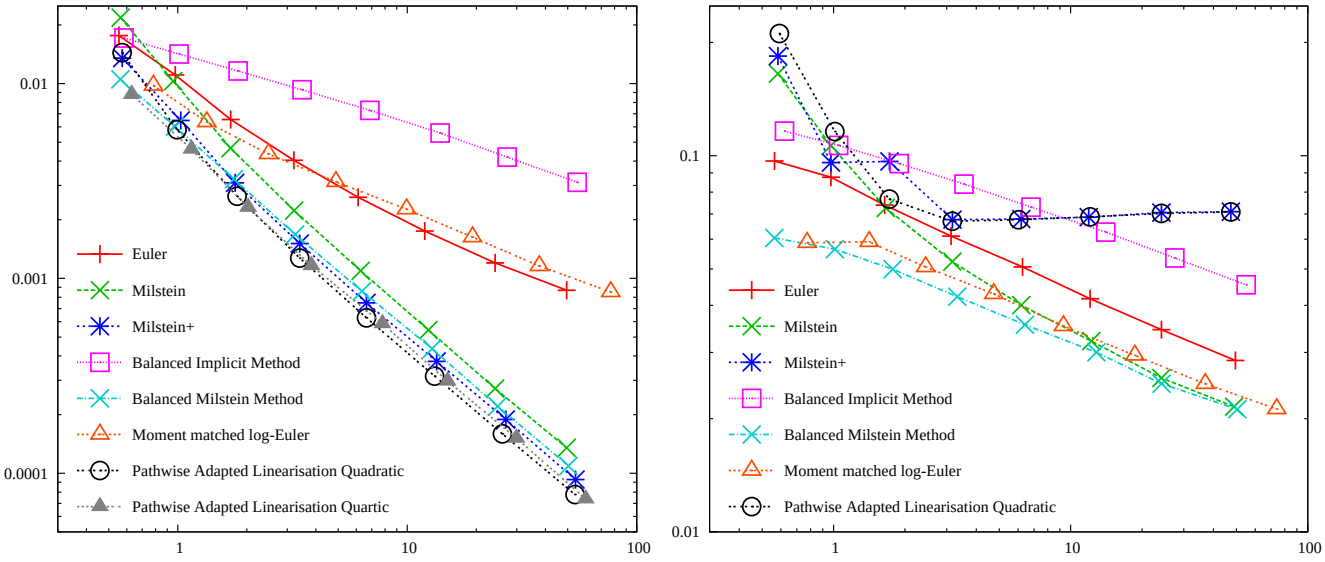

Figure 5: Strong convergence measured by expression (3.15) as a function of CPU time [in msec] averaged over 32767 paths for the mean reverting CEV model (2.2) for $q = 1/2$, $\kappa = 1$, $V_0 = \theta = 1/16$, $T = 1$, $c_0^{\text{BIM}} = 1$, $c_1^{\text{BIM}} = \alpha/\sqrt{x}$, $d_0^{\text{BMM}} = \kappa$, $d_1^{\text{BMM}} = 0$. The number generator was the Sobol' method. (A): $\alpha = 0.2$, $\alpha^2 - 2\kappa\theta = -0.085$; zero is unattainable. (B): $\alpha = 0.8$, $\alpha^2 - 2\kappa\theta = 0.515$; zero is attainable.

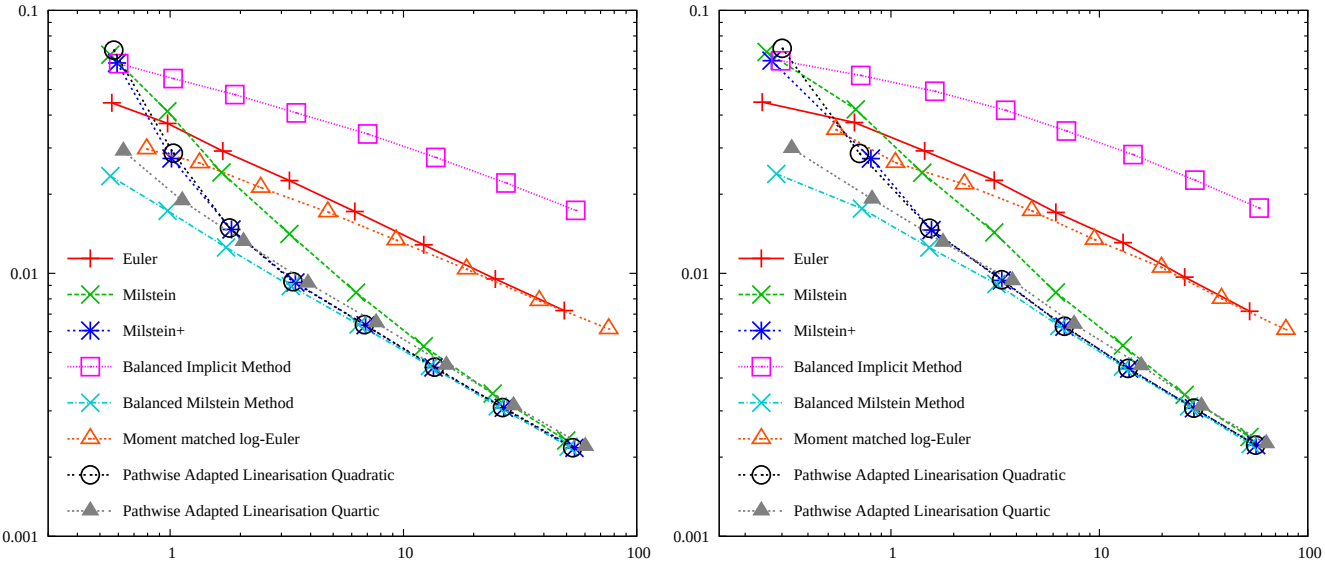

Figure 6: Strong convergence measured by expression (3.15) as a function of CPU time [in msec] averaged over 32767 paths for the mean reverting CEV model (2.2) for $q = 1/2$, $\kappa = 1$, $V_0 = \theta = 1/16$, $\alpha = 0.5$, $\alpha^2 - 2\kappa\theta = 0.125$, zero is attainable, $T = 1$, $c_0^{\text{BIM}} = 1$, $c_1^{\text{BIM}} = \alpha/\sqrt{x}$, $d_0^{\text{BMM}} = \kappa$, $d_1^{\text{BMM}} = 0$. The number generator method was (A) Sobol's and (B) the Mersenne Twister.

The easiest approach for the numerical integration of (4.2) is the Euler-Maruyama scheme

$$\ln S_{t_{n+1}} \;=\; \ln S_{t_n} + \mu\Delta t_n - \tfrac{1}{2}V_{t_n}^{2p}\Delta t_n + V_{t_n}^{p}\Delta W_n \;. \tag{4.3}$$

This scheme has strong convergence order $1/2$, is very easy to implement, and will be our benchmark for all other methods discussed in the following.

An alternative is of course the two-dimensional Milstein scheme (see Appendix A.3) which has strong convergence order 1. It requires the simulation of the double Wiener integral

$$\tilde{I}_{(2,1)}(t_0,t) = \int_{t_0}^{t}\int_{t_0}^{s} \mathrm{d}\tilde{W}_2(u)\mathrm{d}\tilde{W}_1(s) \tag{4.4}$$

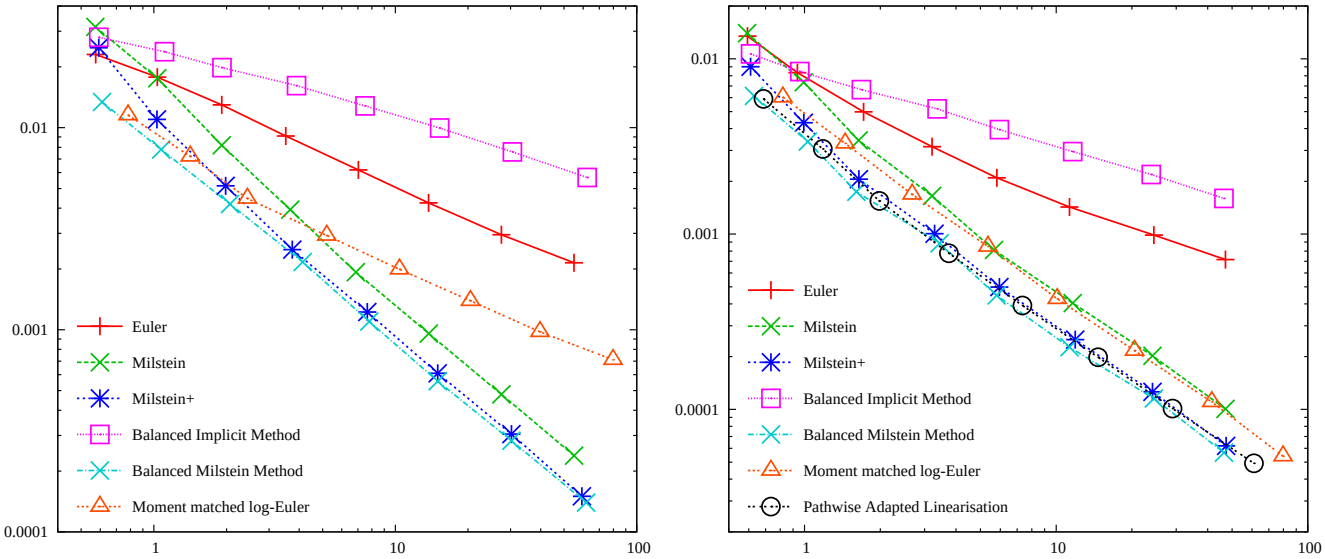

Figure 7: Strong convergence measured by expression (3.15) as a function of CPU time [in msec] averaged over 32767 paths for the mean reverting CEV model (2.2) for $\kappa = 1$, $V_0 = \theta = 0.0625 = 1/16$, $T = 1$. The number generator was the Sobol' method. (A): $q = 3/4$, $c_0^{\text{BIM}} = 1$, $c_1^{\text{BIM}} = \alpha|x|^{-1/4}$, $d_0^{\text{BMM}} = \kappa/2 + 3/8\,\alpha^2/\sqrt{x}$, $d_1^{\text{BMM}} = 0$. (B): $q = 1$, $c_0^{\text{BIM}} = 1$,$c_1^{\text{BIM}} = \alpha$, $d_0^{\text{BMM}} = 1/2\left(\kappa + \alpha^2\right)$, $d_1^{\text{BMM}} = 0$.

for two uncorrelated standard Wiener processes $\tilde{W}_1$ and $\tilde{W}_2$. The standard approximation for this cross term requires several additional random numbers which we consider undesirable for the same reasons we gave to exclude the adaptive Milstein scheme (3.6). There are, however, approaches [Abe04, GL94] to avoid the drawing of many extra random numbers by using the relation of this integral to the Levy-area [Lév51]

$$A_{(1,2)}(t_0,t) = \int_{t_0}^{t}\int_{t_0}^{s}\left(\mathrm{d}\tilde{W}_1(u)\mathrm{d}\tilde{W}_2(s) - \mathrm{d}\tilde{W}_2(u)\mathrm{d}\tilde{W}_1(s)\right) . \tag{4.5}$$

The idea is to employ

$$\int_{t_0}^{t}\int_{t_0}^{s}\left(\mathrm{d}\tilde{W}_1(u)\mathrm{d}\tilde{W}_2(s) + \mathrm{d}\tilde{W}_2(u)\mathrm{d}\tilde{W}_1(s)\right) = \Delta\tilde{W}_1^{(t_0,t)}\Delta\tilde{W}_2^{(t_0,t)} \tag{4.6}$$

to obtain

$$\tilde{I}_{(2,1)}(t_0,t) = \tfrac{1}{2}\left(\Delta\tilde{W}_1^{(t_0,t)}\Delta\tilde{W}_2^{(t_0,t)} - A_{(1,2)}(t_0,t)\right) . \tag{4.7}$$

The joint density of the Levy-area is known semi-analytically

$$\Psi(a,b,c) = \frac{1}{2\pi^2}\int_0^\infty \frac{x}{\sinh(x)}\mathrm{e}^{\frac{-(b^2+c^2)x}{2\tanh(x)}}\cos(ax)\mathrm{d}x \tag{4.8}$$

with $a = A_{(1,2)}(0,1)$, $b = \Delta\tilde{W}_1^{(0,1)}$ and $c = \Delta\tilde{W}_2^{(0,1)}$. Hence, the simulation of the double integral (4.4) is reduced to the drawing of one additional random number (conditional on $\Delta\tilde{W}_1$ and $\Delta\tilde{W}_2$) from this distribution. Gaines and Lyons [GL94] used a modification of Marsaglia's rectangle-wedge-tail method (see [MAP76, MMB64]) to draw from (4.8) which works well for small stepsizes Δt_n. We are, however, interested in methods that also work well for moderately large step sizes, and are simple in their evaluation analytics in order to be sufficiently fast to be useful for industrial purposes.

In essence, all of the above means that we would like to construct a fast numerical integration scheme without the need for auxiliary random numbers. The formal solution (4.2) requires that we handle two

stochastic integral terms. First, we need to approximate the stochastic part of the drift

$$\int_{t_0}^{t} V_s^{2p} \mathrm{d}s \ , \tag{4.9}$$

and secondly, we have to simulate the diffusion term

$$\int_{t_0}^{t} V_s^{p} \mathrm{d}W_s \ . \tag{4.10}$$

For both parts we make intensive use of the Itô-Taylor expansion of the process followed by the m-th power of V_s,

$$V_s^m \ = \ V_{t_0}^m + \int_{t_0}^{s} m V_u^{m-1} b(V_u) \mathrm{d}Z_u + \int_{t_0}^{s} \left(m V_u^{m-1} a(V_u) + \tfrac{1}{2} m(m-1) V_u^{m-2} b^2(V_u) \right) \mathrm{d}u \ , \tag{4.11}$$

with positive exponent m, for any $s \in [t_0, t]$. The term that dominates the overall scheme's convergence is the Wiener integral over $\mathrm{d}Z_u$.

4.1 Interpolation of the drift term (4.9)

A simple way to improve the approximation of the drift integral somewhat is

$$\int_{t_n}^{t_{n+1}} V_s^{2p} \mathrm{d}s \ \approx \ \tfrac{1}{2} \left(V_{t_n}^{2p} + V_{t_{n+1}}^{2p} \right) \Delta t_n \ , \qquad \Delta t_n = (t_{n+1} - t_n) \tag{4.12}$$

which gives us

$$\ln S_{t_{n+1}} \ = \ \ln S_{t_n} + \mu \Delta t_n - \tfrac{1}{4} \left(V_{t_n}^{2p} + V_{t_{n+1}}^{2p} \right) \Delta t_n + V_{t_n}^{p} \Delta W_n \ . \tag{4.13}$$

This *Drift interpolation* scheme comprises practically no additional numerical effort due to the fact that we already know the whole path of the volatility V_{t_i}. Unfortunately, a pure drift interpolation has only a minor impact on the strong approximation quality. Moreover, having a closer look at figure 8, it seems that the *Drift interpolation* method is inferior to the standard log-Euler scheme (4.3). Nevertheless, this approximation has some side effects of benefit for applications that are not fully strongly path dependent whence we discuss it in more detail.

In order to analyse the Drift interpolation scheme (4.13), we start with the Itô-Taylor expansion of the integral of the $2p$-th power of stochastic volatility by setting $m = 2p$ in equation (4.11) to obtain

$$\begin{aligned} \int_{t_n}^{t_{n+1}} V_s^{2p} \mathrm{d}s \ \approx \ & \underbrace{V_{t_n}^{2p} \int_{t_n}^{t_{n+1}} \mathrm{d}s}_{\text{Euler}} + \underbrace{\left(2p V_{t_n}^{2p-1} b_n \right) \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}Z_u \mathrm{d}s}_{\text{First remainder term: } R_1} \\ & + \underbrace{\left(2p V_{t_n}^{2p-1} a_n + p(2p-1) V_{t_n}^{2p-2} b_n^2 \right) \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}u \, \mathrm{d}s}_{\text{Second remainder term: } R_2} \end{aligned} \tag{4.14}$$

with $\Delta t_n := (t_{n+1} - t_n)$, $a_n := a(V_{t_n})$, and $b_n := b(V_{t_n})$. In comparison, the Itô-Taylor expansion of the drift-interpolation scheme (4.12) leads to

$$\begin{aligned} \tfrac{1}{2}\Delta t_n \cdot \left(V_{t_n}^{2p} + V_{t_{n+1}}^{2p}\right) \;\approx\; \tfrac{1}{2}\Delta t_n \cdot \Bigg(& V_{t_n}^{2p} + V_{t_n}^{2p} + \left(2pV_{t_n}^{2p-1} b_n\right) \Delta Z_n \\ & + \left[2pV_{t_n}^{2p-1} a_n + p(2p-1)V_{t_n}^{2p-2} b_n^2\right] \Delta t_n \Bigg) \,. \end{aligned} \tag{4.15}$$

This means that the leading order terms of the local approximation error incurred by the drift interpolation scheme are

$$\begin{aligned} f_{t_n} \;&:=\; \int_{t_n}^{t_{n+1}} V_s^{2p} \mathrm{d}s - \tfrac{1}{2}\left(V_{t_n}^{2p} + V_{t_{n+1}}^{2p}\right)\Delta t_n & (4.16) \\ &= 2pV_{t_n}^{2p-1} b_n \int_{t_n}^{t_{n+1}} \left(\int_{t_n}^{s} \mathrm{d}Z_u - \tfrac{1}{2}\int_{t_n}^{t_{n+1}} \mathrm{d}Z_u\right) \mathrm{d}s \\ &= 2pV_{t_n}^{2p-1} b_n \int_{t_n}^{t_{n+1}} \left(\tfrac{1}{2}\int_{t_n}^{t_{n+1}} \mathrm{d}u - \int_{t_n}^{s} \mathrm{d}u\right) \mathrm{d}Z_s \,. & (4.17) \end{aligned}$$

Thus, by interpolating the drift, the term on the second line of (4.14) involving the double integral

$$I_{(0,0)}(t_n, t_{n+1}) \;=\; \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}u \, \mathrm{d}s \tag{4.18}$$

is catered for. In expectation, we have the unconditional local mean-approximation error

$$\mathsf{E}[f_{t_n} | \mathcal{F}_0] = \mathcal{O}\left(\Delta t_n^3\right) \,. \tag{4.19}$$

In order to analyse the relation between local and global convergence properties, we assume that the integration interval $[0, T]$ is discretised in N steps, $0 < t_1 < \ldots < t_{N-1} < t_N = T$ with stepsize $\Delta t = \frac{T}{N}$. Let $X_{t_i,x}(t_{i+1})$ be the numerical approximation at t_{i+1} starting at time t_i at point x and let $Y_{t_i,x}(t_{i+1})$ be the analytical solution of the stochastic differential equation starting at (t_i, x). Furthermore, we already know the local mean-approximation errors for $i = 0, \ldots, N-1$,

$$\mathsf{E}\left[\left|X_{t_i,Y_i}(t_{i+1}) - Y_{t_i,Y_i}(t_{i+1})\right| \,\middle|\, \mathcal{F}_{t_i}\right] = \mathcal{O}\left(\Delta t_n^3\right) . \tag{4.20}$$

Next we consider the global mean-approximation error

$$\begin{aligned} \left|\mathsf{E}[X_{0,X_0}(T)] - \mathsf{E}[Y_{0,X_0}(T)]\right| \;&=\; \left|\mathsf{E}[X_{0,X_0}(T) - Y_{0,X_0}(T)]\right| & (4.21) \\ &= \left|\mathsf{E}\left[X_{0,X_0}(t_{N-1}) - Y_{0,X_0}(t_{N-1}) + \mathcal{O}\left(\Delta t^3\right)\right]\right| \\ &= \left|\mathsf{E}\left[X_{0,X_0}(t_1) - Y_{0,X_0}(t_1) + (N-1)\cdot\mathcal{O}\left(\Delta t^3\right)\right]\right| \\ &= N \cdot \mathcal{O}\left(\Delta t^3\right) \\ &= \mathcal{O}(\Delta t^2) \,. & (4.22) \end{aligned}$$

This means, the use of the drift interpolation term $\frac{1}{2}\left(V_{t_n}^{2p} + V_{t_{n+1}}^{2p}\right)\Delta t_n$ instead of the straightforward Euler scheme term $V_{t_n}^{2p}\Delta t_n$ improves the global mean-approximation order of convergence. Alas, it is not possible to improve the *global weak*[4] order of convergence in the two-dimensional case without generating additional random numbers. Nevertheless, the interpolation of the drift leads to a higher global

[4]The global weak order of convergence is defined by $|\mathsf{E}[g(X_{0,X_0}(T))] - \mathsf{E}[g(Y_{0,X_0}(T))]|$ with g being a sufficiently smooth test-function. One can find the multidimensional second order weak Taylor approximation scheme in section 14.2 of [KP99].

mean-convergence order (4.22) which may be of benefit when the simulation target is the valuation of plain-vanilla or weakly path dependent options, and this issue will be the subject of future research.

Having analysed the approximation quality of the term governed by $I_{(0,0)}$ in (4.13), we now turn our attention to the local estimation error induced by the handling of the double Wiener integral

$$I_{(2,0)}(t_n, t_{n+1}) \;=\; \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}Z_u \mathrm{d}s \tag{4.23}$$

which can be simulated by the aid of our knowledge of the distribution $I_{(2,0)}(t_n, t_{n+1})$:

$$I_{(2,0)}(t_n, t_{n+1}) \;\sim\; \frac{1}{2}\Delta Z_n \cdot \Delta t_n + \frac{1}{2\sqrt{3}}\epsilon \cdot \Delta t_n, \qquad \text{with } \epsilon \sim \mathcal{N}(0, \Delta t_n)\ . \tag{4.24}$$

Sampling $I_{(2,0)}$ exactly would thus require the generation of an additional random number ϵ for each step. In analogy to the reasoning leading up to the approximation (A.14) which is at the basis of the Milstein+ scheme in appendix A.2, we argue that

$$I_{(2,0)}(t_n, t_{n+1}) \;\rightsquigarrow\; \frac{1}{2}\Delta Z_n \cdot \Delta t_n \tag{4.25}$$

is, conditional on our knowledge of the simulated Wiener path that drives the volatility process, or, more formally, conditional on the σ-algebra $\mathcal{P}_N^2$ generated by the increments

$$\Delta Z_0 = Z_1 - Z_0, \qquad \Delta Z_1 = Z_2 - Z_1, \qquad \ldots, \qquad \Delta Z_{N-1} = Z_N - Z_{N-1}\ , \tag{4.26}$$

the best approximation attainable without resorting to additional sources of (pseudo-)randomness. Applying the approximation (4.25) to the term R_1 in (4.14) leads us to precisely the corresponding term in the expansion (4.15) (last term on the first line) of the drift interpolation scheme, and hence the scheme (4.13) also aids with respect to the influences of the term $I_{(2,0)}(t_n, t_{n+1})$.

The conditional expectation of the local approximation error (4.16) of the scheme (4.13) *conditional on knowing the full path for* Z is thus of order

$$\mathsf{E}\left[f_{t_n} | \mathcal{P}_N^2\right] = \mathcal{O}\left(\Delta Z_n^2 \cdot \Delta t_n\right) + \mathcal{O}\left(\Delta t_n^3\right). \tag{4.27}$$

The quality of this path-conditional local approximation error is not visible in error measures designed to show the strong convergence behaviour of the integration scheme. However, it is likely to be of benefit for the calculation of expectations that do not depend strongly on the fine structure of simulated paths, but on the approximation quality of the distribution of the underlying variable at the terminal time horizon of the simulation.

Another aspect of the drift interpolation scheme (4.13) is that it reduces the local mean-square error

$$\begin{aligned}
\mathsf{E}\left[f_{t_n}^2 | \mathcal{F}_{t_n}\right] &= \left(2pV_{t_n}^{2p-1} b_n\right)^2 \mathsf{E}\left[\left(\int_{t_n}^{t_{n+1}} \left({}^{\Delta t_n}\!/_2 - (s - t_n)\right) \mathrm{d}Z_s\right)^2\right] && (4.28)\\
&= \left(2pV_{t_n}^{2p-1} b_n\right)^2 \int_{t_n}^{t_{n+1}} \left({}^{\Delta t_n}\!/_2 - (s - t_n)\right)^2 \mathrm{d}s && (4.29)\\
&= \left(2pV_{t_n}^{2p-1} b_n\right)^2 \frac{1}{12}\Delta t_n^3 && (4.30)
\end{aligned}$$

compared with the mean-square error of the first remainder term R_1 in (4.14) of the Euler scheme

$$\begin{aligned}
\mathsf{E}\left[(R_1)^2 | \mathcal{F}_{t_n}\right] &= \left(2pV_{t_n}^{2p-1} b_n\right)^2 \mathsf{E}\left[\left(\int_{t_n}^{t_{n+1}} s\, \mathrm{d}Z_s\right)^2\right] && (4.31)\\
&= \left(2pV_{t_n}^{2p-1} b_n\right)^2 \frac{1}{3}\Delta t_n^3\ . && (4.32)
\end{aligned}$$

In summary, the interpolation of the drift given by the scheme (4.13) effectively improves the numerical integration by fully representing terms governed by $I_{(0,0)}$ in the Itô-Taylor expansion of the formal solution, and by improving the approximation for the term governed by $I_{(2,0)}$. We could not really expect to enhance the global strong convergence order induced by the drift term (4.9) without the drawing of additional random numbers. Still, with nearly no extra computational effort one can improve, at least theoretically, over the conventional Euler scheme. Specifically, we are not completely erasing the leading error term of the Euler scheme which is of order $\mathcal{O}(\Delta Z \cdot \Delta t)$. However, by using approximation (4.15), conditional on any one given path in Z, we are able to remove the leading order bias term which is of order $\mathcal{O}(\Delta Z \cdot \Delta t)$. Effectively, the drift interpolation scheme (4.13) simply reduces the absolute value of the coefficient of the lowest strong convergence order error term.

4.2 Mixed interpolation of the diffusion term (4.10)

A suitable approximation of the diffusion is a little bit more difficult than the integration of the drift. The first idea might be to use

$$\int_{t_n}^{t_{n+1}} V_s^p \mathrm{d}W_s \approx \tfrac{1}{2}\left(V_{t_n}^p + V_{t_{n+1}}^p\right) \Delta W_n \ , \tag{4.33}$$

resulting in

$$\ln S_{t_{n+1}} \ = \ \ln S_{t_n} + \mu \Delta t_n - \tfrac{1}{2} V_{t_n}^{2p} \Delta t_n + \tfrac{1}{2}\left(V_{t_n}^p + V_{t_{n+1}}^p\right) \Delta W_n \tag{4.34}$$

which was a simple interpolation for the drift approximation. Furthermore combining the drift and diffusion interpolation leads to

$$\ln S_{t_{n+1}} \ = \ \ln S_{t_n} + \mu \Delta t_n - \tfrac{1}{4}\left(V_{t_n}^{2p} + V_{t_{n+1}}^{2p}\right) \Delta t_n + \tfrac{1}{2}\left(V_{t_n}^p + V_{t_{n+1}}^p\right) \Delta W_n \ . \tag{4.35}$$

We will denote these schemes as *Diffusion interpolation* (4.34) and *Drift + Diffusion interpolation* (4.35). Considering figure 8 we recognize that these integration schemes are remarkable effective in case of no correlation between the underlying and the stochastic volatility. In contrast, convergence is lost al-

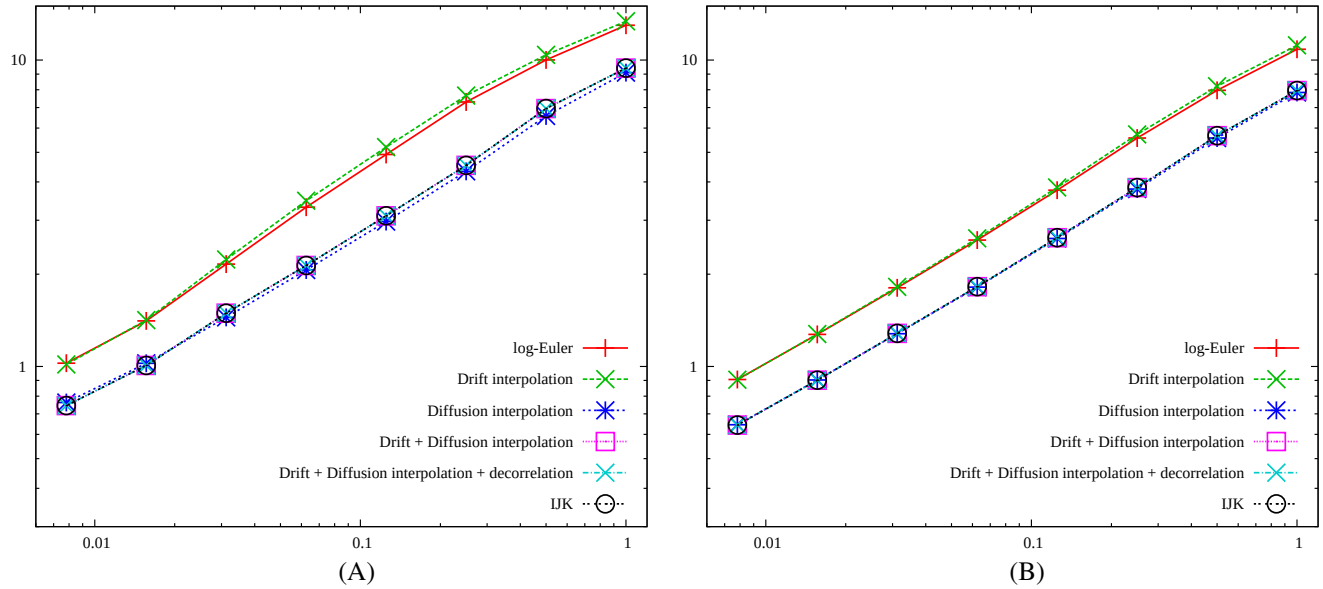

Figure 8: Strong convergence of the financial underlying measured by expression (3.15) averaged over 32767 paths as a function of scheme step size for $T = 1$, $\mu = 0.05$, $S_0 = 100$, $\rho = 0$. The volatility dynamics were given by the (A) exponentially (2.24) and (B) the hyperbolically (2.25) transformed Ornstein-Uhlenbeck process (2.4) with $y_0 = 0$, $\sigma_0 = 1/4$, $\kappa = 1$, and $\alpha = 7/20$.

together for the diffusion interpolation scheme (4.34) when correlation is non-zero as we can see in figures 9 and 10.

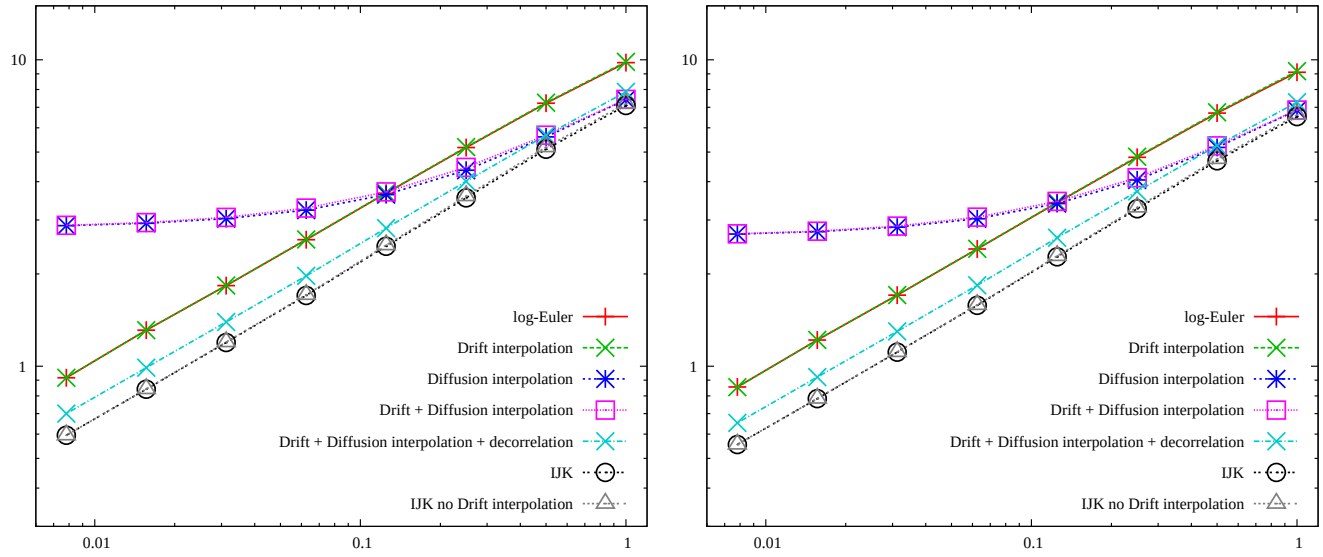

Figure 9: Strong convergence of the financial underlying measured by expression (3.15) averaged over 32767 paths as a function of scheme step size for $T = 1$, $\mu = 0.05$, $S_0 = 100$, $\rho = -2/5$. The volatility dynamics were given by the (A) exponentially (2.24) and (B) the hyperbolically (2.25) transformed Ornstein-Uhlenbeck process (2.4) with $y_0 = 0$, $\sigma_0 = 1/4$, $\kappa = 1$, and $\alpha = 7/20$.

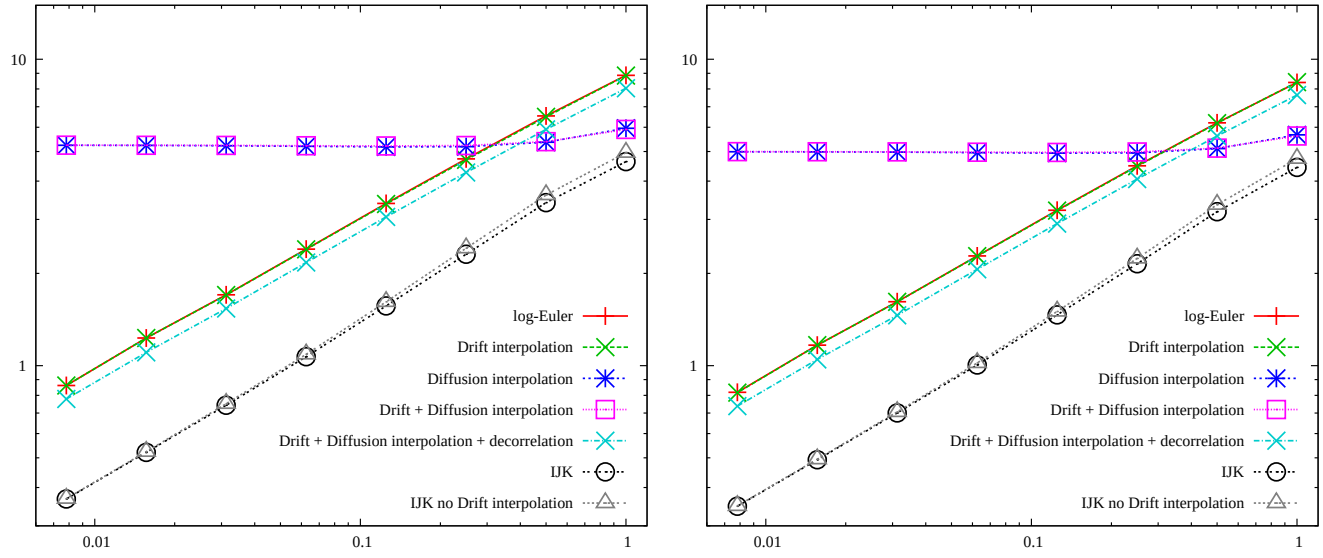

Figure 10: Strong convergence of the financial underlying measured by expression (3.15) averaged over 32767 paths as a function of scheme step size for $T = 1$, $\mu = 0.05$, $S_0 = 100$, $\rho = -4/5$. The volatility dynamics were given by the (A) exponentially (2.24) and (B) the hyperbolically (2.25) transformed Ornstein-Uhlenbeck process (2.4) with $y_0 = 0$, $\sigma_0 = 1/4$, $\kappa = 1$, and $\alpha = 7/20$.

In order to understand the loss of convergence we take a closer look at the diffusion interpolation. The first step is to decompose the correlated Wiener processes into independent components by the aid of the Cholesky decomposition

$$\mathrm{d}W = \rho\,\mathrm{d}\tilde{Z} + \rho'\mathrm{d}\tilde{W}\,, \tag{4.36}$$

$$\mathrm{d}Z = \mathrm{d}\tilde{Z}\,, \tag{4.37}$$

where $\tilde{W}$ and $\tilde{Z}$ are uncorrelated, and

$$\rho' := \sqrt{1-\rho^2}\,. \tag{4.38}$$

This gives us

$$\int_{t_n}^{t_{n+1}} V_s^p \mathrm{d}W_s = \rho' \int_{t_n}^{t_{n+1}} V_s^p \mathrm{d}\tilde{W}_s + \rho \int_{t_n}^{t_{n+1}} V_s^p \mathrm{d}\tilde{Z}_s \ . \tag{4.39}$$

The reason for the loss of convergence is that the volatility process V_s is driven itself by the Wiener process $\tilde{Z}_s$. Thus by using the trapezoidal rule we are not interpreting the stochastic integral in the Itô but in the Stratonovich sense. As a consequence, we are overestimating the influence of the Wiener process $\tilde{Z}_s$. We can circumvent this problem by applying the trapezoidal rule only on the uncorrelated part of the diffusion

$$\rho' \int_{t_n}^{t_{n+1}} V_s^p \mathrm{d}\tilde{W}_s + \rho \int_{t_n}^{t_{n+1}} V_s^p \mathrm{d}\tilde{Z}_s \quad \approx \quad \tfrac{1}{2}\rho' \left(V_{t_n}^p + V_{t_{n+1}}^p\right) \Delta\tilde{W}_n + \rho V_{t_n}^p \Delta\tilde{Z}_n \ , \tag{4.40}$$

which gives us in combination with (4.35)

$$\ln S_{t_{n+1}} \ = \ \ln S_{t_n} + \mu\Delta t_n - \tfrac{1}{4}\left(V_{t_n}^{2p} + V_{t_{n+1}}^{2p}\right)\Delta t_n + \tfrac{1}{2}\left(V_{t_n}^p + V_{t_{n+1}}^p\right)\Delta W_n + \tfrac{1}{2}\left(V_{t_n}^p - V_{t_{n+1}}^p\right)\rho\Delta Z_n \tag{4.41}$$

which we shall refer to as *Drift + Diffusion interpolation + decorrelation* scheme. This scheme not only restores convergence but also improves the approximation quality when correlation is non-zero. To verify these statements we analyse the local approximation error again

$$f_{t_n} \ := \rho' \int_{t_n}^{t_{n+1}} V_s^p \mathrm{d}\tilde{W}_s + \rho \int_{t_n}^{t_{n+1}} V_s^p \mathrm{d}\tilde{Z}_s - \left(\tfrac{1}{2}\rho'\left(V_{t_n}^p + V_{t_{n+1}}^p\right)\Delta\tilde{W}_n + \rho V_{t_n}^p \Delta\tilde{Z}_n\right) \tag{4.42}$$

$$\begin{aligned}
&= \rho' \left(pV_{t_n}^{p-1} b_n \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}\tilde{Z}_u \mathrm{d}\tilde{W}_s + \left[pV_{t_n}^{p-1} a_n + \tfrac{1}{2}p(p-1)V_{t_n}^{p-2} b_n^2\right] \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}u \, \mathrm{d}\tilde{W}_s \right) \\
&\quad + \rho \left(pV_{t_n}^{p-1} b_n \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}\tilde{Z}_u \mathrm{d}Z_s + \left[pV_{t_n}^{p-1} a_n + \tfrac{1}{2}p(p-1)V_{t_n}^{p-2} b_n^2\right] \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}u \, \mathrm{d}\tilde{Z}_s \right) \\
&\quad - \tfrac{1}{2}\rho' \left(pV_{t_n}^{p-1} b_n \cdot \Delta\tilde{Z}_n + \left[pV_{t_n}^{p-1} a_n + \tfrac{1}{2}p(p-1)V_{t_n}^{p-2} b_n^2\right] \cdot \Delta t_n \right) \cdot \Delta\tilde{W}_n \\
&= \rho' \Bigg(\underbrace{pV_{t_n}^{p-1} b_n \int_{t_n}^{t_{n+1}} \left((\tilde{Z}_s - \tilde{Z}_{t_n}) - \tfrac{1}{2}\Delta\tilde{Z}_{t_n} \right) \mathrm{d}\tilde{W}_s}_{f_{t_n,1}} \\
&\quad + \underbrace{\left(pV_{t_n}^{p-1} a_n + \tfrac{1}{2}p(p-1)V_{t_n}^{p-2} b_n^2\right) \int_{t_n}^{t_{n+1}} \left((s - t_n) - \tfrac{1}{2}\Delta t_n \right) \mathrm{d}\tilde{W}_s}_{f_{t_n,2}} \Bigg) \\
&\quad + \rho \left(\underbrace{pV_{t_n}^{p-1} b_n \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}\tilde{Z}_u \mathrm{d}\tilde{Z}_s}_{f_{t_n,3}} + \left[pV_{t_n}^{p-1} a_n + \tfrac{1}{2}p(p-1)V_{t_n}^{p-2} b_n^2\right] \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}u \, \mathrm{d}\tilde{Z}_s \right) \ .
\end{aligned} \tag{4.43}$$

In analogy to the interpolation of the drift, the trapezoidal integration rule applied to the uncorrelated part of the Itô integral in (4.40) leads to a reduced variance for the local truncation errors $f_{t_n,1}$ and $f_{t_n,2}$. Taking the conditional expectation based on the knowledge of our Wiener paths W and Z we obtain

$$\mathsf{E}\left[f_{t_n,1}|\tilde{\mathcal{P}}_N^1,\tilde{\mathcal{P}}_N^2\right]=0, \qquad \text{and} \qquad \mathsf{E}\left[f_{t_n,2}|\tilde{\mathcal{P}}_N^1\right]=0\ , \tag{4.44}$$

where the σ-algebras $\tilde{\mathcal{P}}_N^1$ and $\tilde{\mathcal{P}}_N^2$ are generated by the increments of the Wiener processes $\tilde{W}$ and $\tilde{Z}$. Once again the interpolation is the best estimate based on the knowledge of the paths of $\tilde{W}$ and $\tilde{Z}$. Especially in the case of low correlation this scheme is remarkable effective. Taking a closer look at the local approximation error (4.43) of the correlated part we recognize that the leading error term is

$$f_{t_n,3}=pV_{t_n}^{p-1}b_n\int_{t_n}^{t_{n+1}}\int_{t_n}^{s}\mathrm{d}\tilde{Z}_u\mathrm{d}\tilde{Z}_s=pV_{t_n}^{p-1}b_n\tfrac{1}{2}\left(\Delta\tilde{Z}_n^2-\Delta t_n\right)\ . \tag{4.45}$$

To make matters even better, we can improve the integration by including this term to our integration scheme. Luckily the double Itô integral $I_{(2,2)}$ does not require additional random numbers. The importance of the inclusion of this term grows with increasing correlation coefficient ρ, unlike the benefit form the (de-correlated) diffusion interpolation (4.41) which diminishes with increasing correlation. For the sake of brevity, we will call this integration scheme based on interpolation of the drift, interpolation of the diffusion term, consideration of decorrelation of the diffusion term, and inclusion of a higher order Milstein term, simply the *IJK* scheme in the following. Its explicit propagation equation is given by

$$\begin{aligned}\ln S_{t_{n+1}} &= \ln S_{t_n}+\mu\Delta t_n-\tfrac{1}{4}\left(V_{t_n}^{2p}+V_{t_{n+1}}^{2p}\right)\Delta t_n+\rho V_{t_n}^{p}\Delta\tilde{Z}_n\\ &\quad+\tfrac{1}{2}\rho'\left(V_{t_n}^{p}+V_{t_{n+1}}^{p}\right)\Delta\tilde{W}_n+\tfrac{1}{2}\rho p V_{t_n}^{p-1}b_n\cdot\left(\Delta\tilde{Z}_n^2-\Delta t_n\right)\end{aligned} \tag{4.46}$$

or, equivalently,

$$\begin{aligned}\ln S_{t_{n+1}} &= \ln S_{t_n}+\mu\Delta t_n-\tfrac{1}{4}\left(V_{t_n}^{2p}+V_{t_{n+1}}^{2p}\right)\Delta t_n+\rho V_{t_n}^{p}\Delta Z_n\\ &\quad+\tfrac{1}{2}\left(V_{t_n}^{p}+V_{t_{n+1}}^{p}\right)\left(\Delta W_n-\rho\Delta Z_n\right)+\tfrac{1}{2}\rho p V_{t_n}^{p-1}b_n\cdot\left(\Delta Z_n^2-\Delta t_n\right)\ .\end{aligned} \tag{4.47}$$

Since the *Drift interpolation* (4.13) scheme was not able to increase the strong approximation quality of the standard log-Euler we also tried the *IJK* method without using a drift interpolation which we denote as *IJK no Drift interpolation*.

In figures 8, 9, and 10 we show all considered approximation procedures in comparison and we see that a combination of drift interpolation, diffusion interpolation allowing for (de-)correlation as given in (4.40), and the addition of the higher order term (4.45) outperforms any of the other approximation schemes. The advantage of the *IJK* scheme is that we get good approximation results for low and high correlations due to the fact that we cover both the dominant error terms for low correlation (4.40) and for high correlation (4.45), and that comparatively little extra computational effort is required. In addition, one can observe that in the case of high correlation, as given in figures 10 and 13, the drift-interpolation is a small but valuable enhancement for the *IJK* scheme particularly for large stepsizes.

Until this point we have only compared integration schemes by looking at the approximation quality as a function of stepsize. In financial applications, however, a scheme is considered *better* if it is *more accurate* and *faster*. It is thus of paramount interest to compare the residual error as a function of calculation time. In figures 11, 12, and 13 we can see that the combination of all speed-ups (drift and diffusion interpolation, decorrelation and additional term) does not affect the computational effort significantly. We also notice that, when simulation of the stochastic volatility process itself is trivial as is the case for the exponential and the hyperbolic volatility processes discussed in section 2.2, the use of the *IJK* scheme provides on average a speed-up of approximately a factor 2

figure	ρ	Exp.	Hyp.
11	0.0	2.0	1.9
12	−0.4	2.2	2.1
13	−0.8	4.4	4.2

Table 1: Average speed-up *IJK* (4.47) compared with *log-Euler* (4.3).

for low correlation and as much as a factor 4 for pronounced negative correlation (see table 1). In this context, it is noteworthy to recall that most market calibrations require a strong negative correlation to reproduce the observable implied volatility skews, which makes the use of the *IJK* scheme particularly attractive.

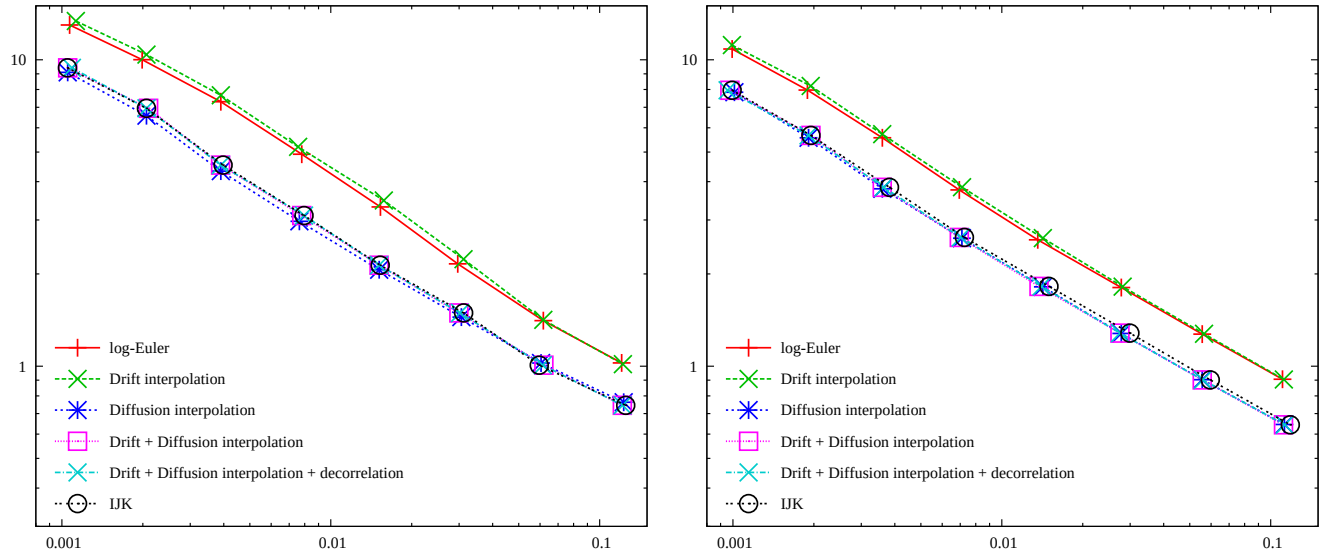

Figure 11: Strong convergence of the financial underlying measured by expression (3.15) averaged over 32767 paths as a function of CPU time [in msec] for $T = 1$, $\mu = 0.05$, $S_0 = 100$, $\rho = 0$. The volatility dynamics were given by the (A) exponentially (2.24) and (B) the hyperbolically (2.25) transformed Ornstein-Uhlenbeck process (2.4) with $y_0 = 0$, $\sigma_0 = 1/4$, $\kappa = 1$, and $\alpha = 7/20$.

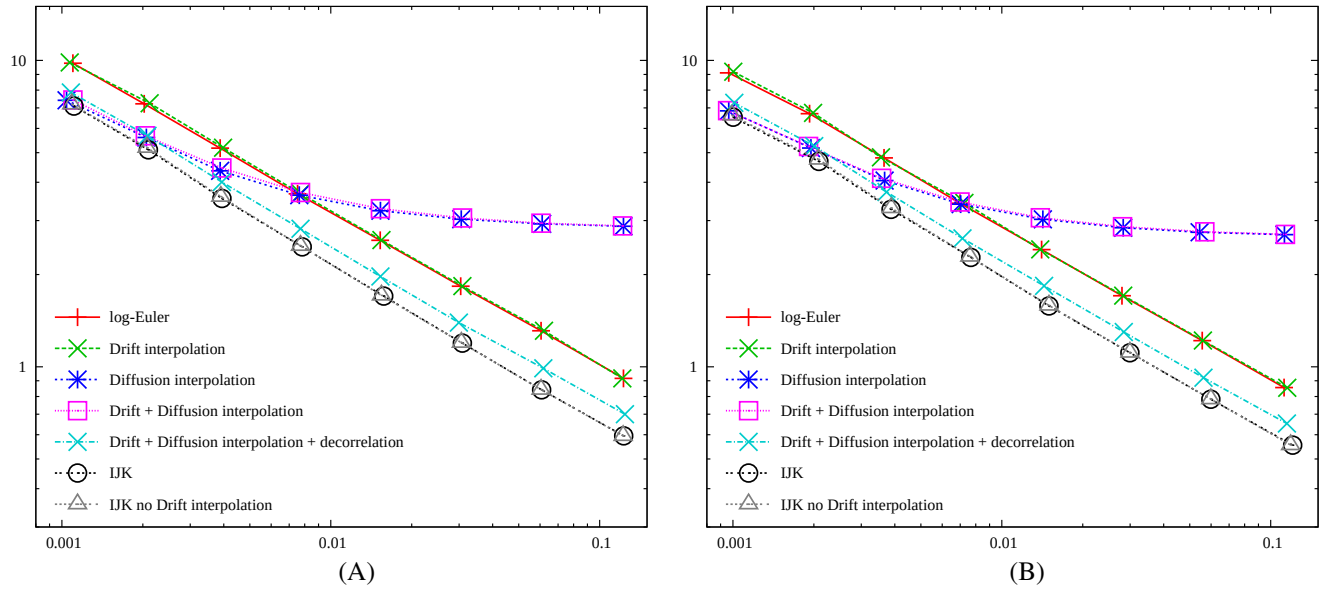

Figure 12: Strong convergence of the financial underlying measured by expression (3.15) averaged over 32767 paths as a function of CPU time [in msec] for $T = 1$, $\mu = 0.05$, $S_0 = 100$, $\rho = -2/5$. The volatility dynamics were given by the (A) exponentially (2.24) and (B) the hyperbolically (2.25) transformed Ornstein-Uhlenbeck process (2.4) with $y_0 = 0$, $\sigma_0 = 1/4$, $\kappa = 1$, and $\alpha = 7/20$.

In section 5, we present the results of further numerical tests for the case when the volatility process itself requires a numerical integration scheme. In particular, we consider the situation where the volatility is given by the mean-reverting CEV process (2.2). For this case, we wish to find the best combination of integration schemes for the stochastic volatility as well as for the financial underlying.

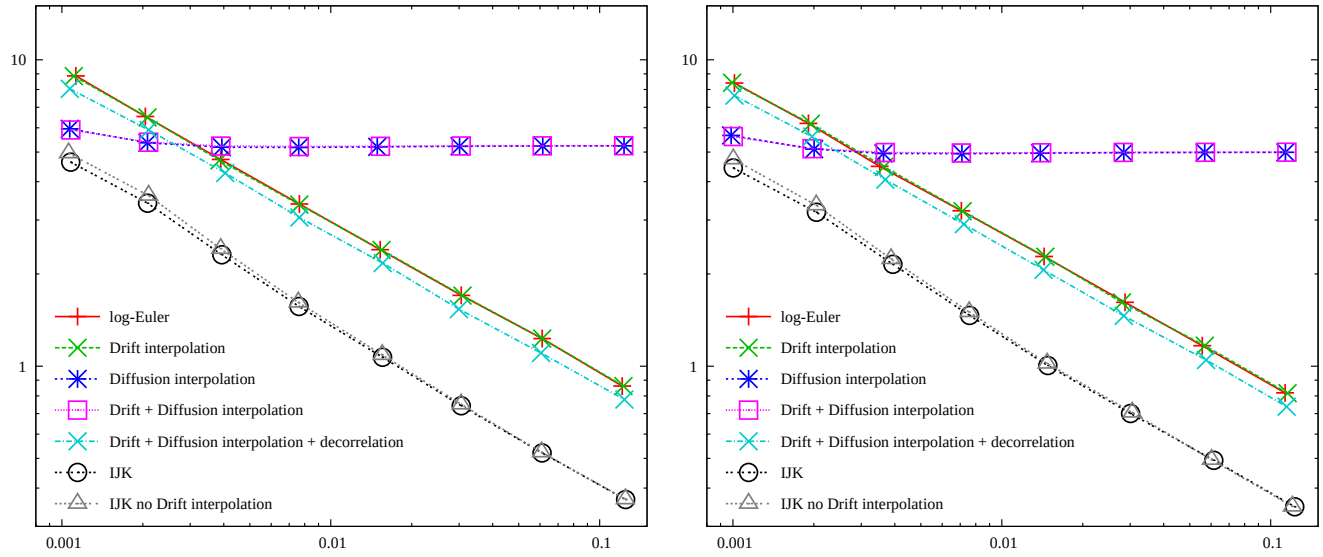

Figure 13: Strong convergence of the financial underlying measured by expression (3.15) averaged over 32767 paths as a function of CPU time [in msec] for $T = 1$, $\mu = 0.05$, $S_0 = 100$, $\rho = -4/5$. The volatility dynamics were given by the (A) exponentially (2.24) and (B) the hyperbolically (2.25) transformed Ornstein-Uhlenbeck process (2.4) with $y_0 = 0$, $\sigma_0 = 1/4$, $\kappa = 1$, and $\alpha = 7/20$.

5 Numerical results for mean-reverting CEV volatility processes

In this section we go one step further as we consider a two-dimensional stochastic volatility model where the stochastic volatility process is given by the mean-reverting CEV process (2.2). We already recognized in section 3 that the numerical results for the integration of a mean-reverting CEV process are sensitive to the size of the diffusion exponent $q \in [1/2, 1]$. Hence we focus on the two extreme choices the Brennan-Schwartz (3.23) and the Cox-Ingersoll-Ross (3.37) equation. In the following we consider four schemes for the integration of stochastic volatility or variance:

1. *Euler* — (3.1) ,
2. *Milstein* — (A.4) ,
3. *BMM* — (3.7) ,
4. *Pathwise Adapted Linearisation* — (3.32) for Brennan-Schwartz and (3.44) for CIR .

We combine these with suitable integration schemes for the whole system. Specifically, we consider

1. *Euler-Maruyama* — (4.3) ,
2. *IJK* — (4.47) .

The Euler scheme was already the benchmark in section 4 where we developed the *IJK* scheme. It is of interest to see if the *IJK* scheme can preserve its advantage even if we have to integrate the stochastic volatility process numerically.

In the following we concentrate on two different test cases. The first one is based on the Brennan-Schwartz equation (3.23) for the modelling of the stochastic volatility. This equation is directly coupled to the underlying with exponent $p = 1/2$

$$\begin{aligned} \mathrm{d}S_t &= \mu S_t \mathrm{d}t + \sqrt{V_t} S_t \mathrm{d}W_t \ , \\ \mathrm{d}V_t &= \kappa \left(\theta - V_t\right) \mathrm{d}t + \alpha V_t \mathrm{d}Z_t \ . \end{aligned} \tag{5.1}$$

The parameter configuration is chosen as follows $S_{t_0} = 100$, $\mu = 0.05$, $V_0 = \theta = 1/16$, $\kappa = 1$, $\alpha = 0.5$ where we present results for different levels of correlation $\rho \in \{0.0, -0.4, -0.8\}$.

As a second benchmark we consider the Heston model

$$\begin{aligned} \mathrm{d}S_t &= \mu S_t \mathrm{d}t + \sqrt{V_t} S_t \mathrm{d}W_t \,, \\ \mathrm{d}V_t &= \kappa\,(\theta - V_t)\,\mathrm{d}t + \alpha\sqrt{V_t}\mathrm{d}Z_t \,, \end{aligned} \tag{5.2}$$

where the parameters are given by $S_{t_0} = 100$, $\mu = 0.05$, $V_0 = \theta = 1/16$, $\kappa = 1$, $\alpha = 0.5$. Again we show results for decreasing correlation $\rho \in \{0.0, -0.4, -0.8\}$.

We see in figures 14–16 that the decisive point for the strong approximation quality is the choice of the integration scheme *IJK* as the integration of the stochastic volatility has just a minor impact on the numerical results. In accordance with the numerical results of the last section, we can observe that the *IJK* scheme is at its most impressive when dealing with high correlation as in figure 16. In fact for high negative correlation, the approximation efficiency of the *IJK* scheme in comparison to conventional Euler & log-Euler methods appears to be even greater in the case when the volatility process itself requires numerical integration considering that the speed gain appears to be approximately a factor 5 in figure 16.

figure	ρ	(5.1)	(5.2)
14	0.0	2.1	2.6
15	-0.4	2.5	2.9
16	-0.8	4.6	4.5

Table 2: Average speed-up *BMM* & *IJK* (4.47) compared with *Euler* & *log-Euler* (4.3).

Nonetheless, even if we can neglect the influence of the numerical integration of the stochastic volatility process on the strong convergence behaviour of the underlying, the details of the integration of the stochastic volatility process become important when pricing derivatives that are sensitive to the dynamics of the volatility. In that case, the results of section 3 can give guidance in the selection of the integration scheme for the mean-reverting CEV process. In any case, one should be aware of the fact that an unstable integration of the stochastic volatility can crash the integration of the whole system in the sense that the occurrence of spurious paths where variance crosses over to the negative domain can spoil the convergence behaviour irrecoverably as we saw in figure 5 (B) for the Milstein+ scheme. On that note, we have a closer look at figure 16 (B) where we see that the convergence behaviour of the *Milstein-IJK* scheme seems somewhat unexpected as the approximation error does not decrease when halving the stepsize from $\Delta t = 1$ to $\Delta t = 1/2$ even though this integration scheme is competitive to the *BMM-IJK* scheme for small stepsizes. The explanation for this is surprisingly simple. In table 3 we compare the percentage of non-positive paths for the integration of the stochastic volatility where we do not count those paths becoming non-positive in the final integration step[5]. With this counting convention, no non-positive paths occur for $\Delta t = 1$ as we only have to take a single step. In comparison, for $\Delta t = 1/2$ we obtain the highest number of non-positive paths for the Milstein scheme which explains the bump in the convergence plot of *Milstein-IJK* in figure 16 (B). Thus, even in this very simple case of estimating the strong convergence error of the financial underlying, an appropriate integration scheme for the stochastic volatility process is key to guaranteeing a stable approximation.

6 Conclusion

In this article, we discussed various Monte-Carlo approximation schemes for stochastic volatility diffusion models. Our main focus was on the strong convergence behaviour as an indicator for the valuation of path dependent derivatives. In order to maintain the ability to apply exogenous variance reduction techniques such as low discrepancy numbers, importance sampling, and others [Jäc02, Gla03], with ease, we restricted our research to methods that effectively require only two simulated uniform variates

Δt	*Euler*	*Milstein*	*BMM*
2^0	0 %	0 %	0 %
2^{-1}	23.9 %	33.2 %	0 %
2^{-2}	37.5 %	15.7 %	0 %
2^{-3}	43.8 %	0.1 %	0 %
2^{-4}	46.7 %	0 %	0 %

Table 3: Number of non-positive stochastic volatility paths in figure 16 (B).

[5] It is of minor importance if one path becomes negative or zero in the last integration step as we do not have to use the final value as a starting point for the next integration step.

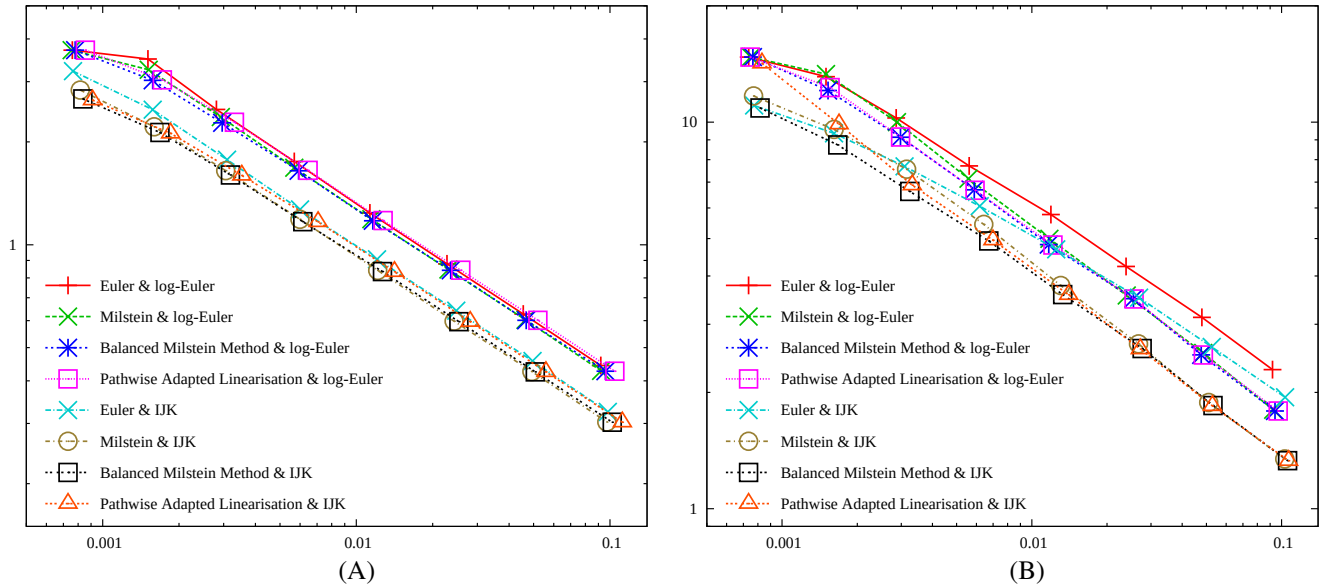

Figure 14: Strong convergence measured by expression (3.15) as a function of CPU time [in msec] averaged over 32767 paths for (A) model (5.1) and (B) for model (5.2). The number generator method was Sobol's. Correlation: $\rho = 0$.

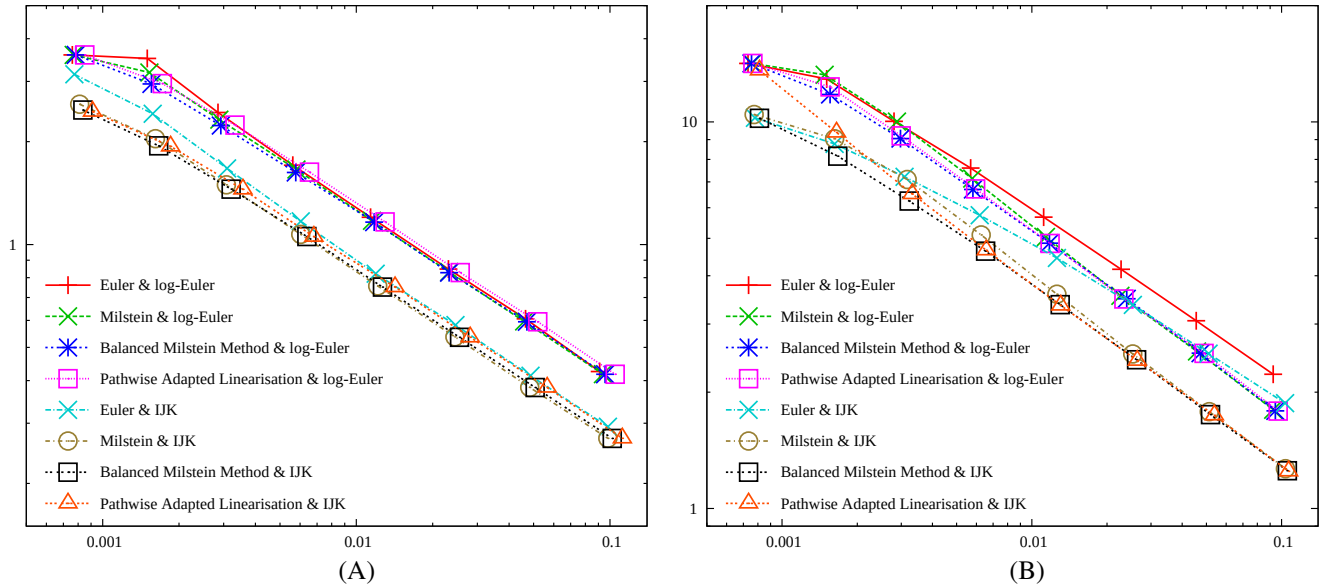

Figure 15: Strong convergence measured by expression (3.15) as a function of CPU time [in msec] averaged over 32767 paths for (A) model (5.1) and (B) for model (5.2). The number generator method was Sobol's. Correlation: $\rho = -0.4$

per step in the time discretisation of the volatility-underlying evolution pair. Given this self-imposed constraint, we attempted to exploit all information available from the simulated primary standard Wiener diffusion process pair, and to adapt the integration scheme of the volatility and financial underlying process as much as possible to each simulated primary process path pair. Whilst we had to realise that within our scope we are limited to improving simulation results not by increasing the convergence *order*, but mainly by decreasing the magnitude of the leading order error term, we found that significant speed gains can be accomplished at surprisingly little expense in numerical effort. In fact, for the stochastic volatility models we examined, the observed acceleration in comparison to the standard Euler & log-Euler varies from a factor two for zero correlation to as much as a factor five when correlation is significantly negative as usually required for calibration to market-observable implied volatility profiles.

As part of our investigations, we also introduced a new variation of stochastic volatility models, namely the *hyperbolically transformed Ornstein-Uhlenbeck* process model given by equation (2.25). This model inherits the benefits of Scott's model [Sco87] such as mean reversion and the fact that zero is not attainable, but, like Scott's model, does not provide us with (semi-)closed form analytical

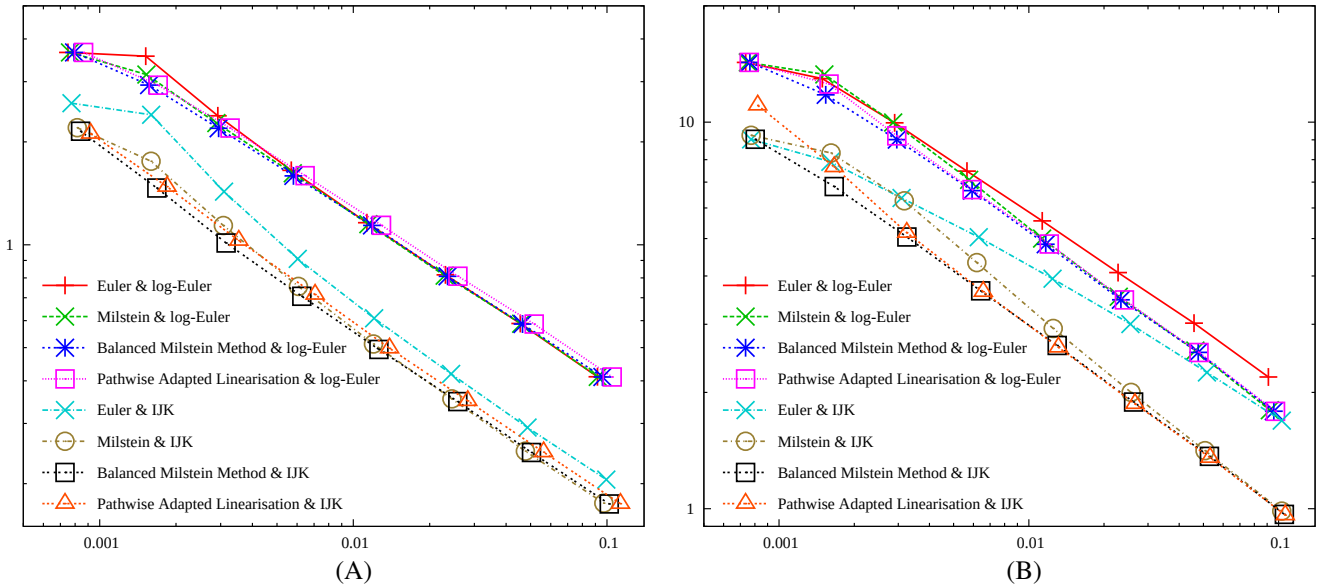

Figure 16: Strong convergence measured by expression (3.15) as a function of CPU time [in msec] averaged over 32767 paths for (A) model (5.1) and (B) for model (5.2). The number generator method was Sobol's. Correlation: $\rho = -0.8$

solutions for the density or characteristic function of the distribution of the underlying, or plain-vanilla option prices. It does, however, avoid the issues raised in [AP04] regarding the explosion of moments etc. due to the fact that the tails of its distribution, both towards low and towards high volatility levels, are significantly *thinner* than those of the Scott model. We examined the behaviour of this model as part of our research because we believe that advanced Monte Carlo integration schemes, in combination with modern variance reduction methods, in conjunction with ever increasing computer power, will make the use of stochastic volatility models that are not readily amenable for convenient plain-vanilla option pricing formulæ an industrially viable possibility, even at the point where model parameters or parameter term structures are calibrated to market observable plain vanilla option prices. This latter conjecture will be the subject of future research.

As for currently favoured stochastic volatility models such as Heston's [Hes93], whose instantaneous variance is driven by the Cox-Ingersoll-Ross process, we found in our numerical experiments that the use of *pathwise adapted linearisations* of the driving Wiener process introduced in section 3.1, with approximate analytical solutions along the path, can provide numerical integration improvements for small to moderate values of α. However, from our set of investigated methods, the clear overall favourite for the integration of the Cox-Ingersoll-Ross process is the Balanced Milstein method since it performs as efficiently as the CIR-specific pathwise adapted linearisation based expansions, but remains stable for all parameter values. The Balanced Milstein method is thus our method of choice for the numerical integration of the CIR process, both inside stochastic volatility model applications and otherwise.

Finally, it remains to be stated that the main focus of our research, namely the investigation of integration schemes for stochastic volatility models that combine an essentially geometric Brownian motion process with a secondary source of noise influencing the volatility term of the former, resulted in the method we named *IJK* scheme given by equation (4.47). The purpose of this method is not to show the highest possible convergence order as a function of the step size. Instead, we endeavoured to find a simple and robust method that, without the need for adaptive refinement, or additional random numbers per step, is as efficient as possible. In other words, we were looking for the most efficient method that is robust and yet essentially as simple as the standard Euler-Maruyama algorithm. The *IJK* scheme is, as a consequence, not superior in convergence order, but excels over other equally simple methods by showing a superior convergence behaviour that is faster by a multiplicative factor. The *IJK* method is as easy to implement as the standard (log-)Euler scheme, and can thus be used as a so-called *drop-in replacement* for the (log-)Euler scheme since it requires no additional random numbers or other convergence acceleration aids.

A Milstein schemes

A.1 The one-dimensional Milstein method

The Milstein scheme [Mil74] for the stochastic differential equation

$$\mathrm{d}x \;=\; a(x)\mathrm{d}t + b(x)\mathrm{d}W \tag{A.1}$$

can be derived from the Itô-Taylor expansion [KP99, equation 5.5.4]

$$\begin{aligned} x_t \;&= x_0 + a_0 \int_0^t \mathrm{d}s + b_0 \int_0^t \mathrm{d}W_s + b_0' b_0 \int_0^t \int_0^s \mathrm{d}W_u \,\mathrm{d}W_s \\ &\quad + a_0' b_0 \int_0^t \int_0^s \mathrm{d}W_u \,\mathrm{d}s + \left(a_0 b_0' + \tfrac{1}{2} b_0'' b_0^2\right) \int_0^t \int_0^s \mathrm{d}u \,\mathrm{d}W_s + \left(b_0 \, {b_0'}^2 + b_0'' b_0^2\right) \int_0^t \int_0^s \int_0^u \mathrm{d}W_r \,\mathrm{d}W_u \,\mathrm{d}W_s \\ &\quad + \mathcal{O}(t^2) \end{aligned} \tag{A.2}$$

with $a_0 = a(x_0)$, etc. Retaining terms up to order $\mathcal{O}(t)$, and evaluating the integral

$$I_{(1,1)}(t) \;:=\; \int_0^t \int_0^s \mathrm{d}W_u \,\mathrm{d}W_s \;=\; \tfrac{1}{2}\left(W_t^2 - t\right) \;, \tag{A.3}$$

assuming (without loss of generality) that $W_0 = 0$, we obtain the one-dimensional Milstein scheme

$$x_{t_{n+1}} \;=\; x_{t_n} + a(x_{t_n})\Delta t_n + b(x_{t_n})\Delta W_n + \tfrac{1}{2} b'(x_{t_n}) b(x_{t_n}) \left(\Delta W_n^2 - \Delta t_n\right) \;. \tag{A.4}$$

Applied to equation (2.2), we obtain

$$V_{t_{n+1}}^{\text{Milstein}} \;=\; V_{t_n} + \kappa(\theta - V_{t_n})\Delta t_n + \alpha V_{t_n}^q \Delta Z_n + \tfrac{1}{2}\alpha^2 q V_{t_n}^{2q-1} \left(\Delta Z_n^2 - \Delta t_n\right) \;. \tag{A.5}$$

A.2 The Milstein+ scheme

All of the explicitly given integral terms on the second line of (A.2)

$$I_{(1,0)}(t) \;=\; \int_0^t \int_0^s \mathrm{d}W_u \,\mathrm{d}s \tag{A.6}$$

$$I_{(0,1)}(t) \;=\; \int_0^t \int_0^s \mathrm{d}u \,\mathrm{d}W_s \;=\; W_t \cdot t - I_{(1,0)}(t) \tag{A.7}$$

$$I_{(1,1,1)}(t) \;=\; \int_0^t \int_0^s \int_0^u \mathrm{d}W_r \,\mathrm{d}W_u \,\mathrm{d}W_s \;=\; \tfrac{1}{6} W_t^3 - \tfrac{1}{2} W_t \cdot t \tag{A.8}$$

can be expressed in terms of the primary constituents W_t and $I_{(1,0)}(t)$. The two-dimensional distribution of the random numbers W_t and $I_{(1,0)}(t)$ is given by a bivariate Gaussian law with mean $(0, 0)$ and covariance matrix

$$\begin{pmatrix} t & t^2/2 \\ t^2/2 & t^3/3 \end{pmatrix} \;. \tag{A.9}$$

This means,

$$I_{(1,0)}(t) \;\sim\; \tfrac{1}{2} \cdot W_t \cdot t + \tfrac{1}{2\sqrt{3}} \cdot t^{3/2} y \tag{A.10}$$

with $y \sim \mathcal{N}(0,1)$, and thus the distribution of x_t is given by

$$\begin{aligned} x_t \;\sim\; & x_0 + a_0 \cdot t + b_0 \cdot W_t + \tfrac{1}{2} b_0' b_0 \left(W_t^2 - t\right) \\ & + \tfrac{1}{6}\left(b_0 \, {b_0'}^2 + b_0'' b_0^2\right) \cdot W_t^3 + \left(a_0 b_0' - \tfrac{1}{2} b_0 \, {b_0'}^2\right) \cdot W_t \cdot t \\ & + \left(a_0' b_0 - a_0 b_0' - \tfrac{1}{2} b_0'' b_0^2\right) \cdot \left[\tfrac{1}{2} \cdot W_t \cdot t + \tfrac{1}{2\sqrt{3}} \cdot t^{3/2} y\right] + \mathcal{O}(t^2) \; . \end{aligned} \tag{A.11}$$

Hence, in order to fully cater for all terms to order $\mathcal{O}(t^{3/2})$, an extra source of randomness is required. It can also be seen that, conditional on a given value for W_t, the terms of order $\mathcal{O}(t^{3/2})$ have non-zero expectation. This means, that, not accounting for the terms on the second and third line of (A.11), i.e. the terms not present in the standard Milstein scheme, introduces a conditional bias given by

$$\tfrac{1}{6}\left[b_0 \, {b_0'}^2 + b_0'' b_0^2\right] \cdot W_t^3 + \tfrac{1}{2}\left[a_0' b_0 + a_0 b_0' - b_0 \, {b_0'}^2 - \tfrac{1}{2} b_0'' b_0^2\right] \cdot W_t \cdot t \; . \tag{A.12}$$

This bias can be corrected by simply adding these terms to the Milstein scheme which gives us

$$\begin{aligned} x_{t_{n+1}} \;=\; & x_{t_n} + a(x_{t_n})\Delta t_n + b(x_{t_n})\Delta W_n + \tfrac{1}{2} b'(x_{t_n}) b(x_{t_n}) \left(\Delta W_n^2 - \Delta t_n\right) \\ & + \tfrac{1}{6}\left[b(x_{t_n}) \, b'(x_{t_n})^2 + b''(x_{t_n}) b(x_{t_n})^2\right] \Delta W_n^3 \\ & + \tfrac{1}{2}\left[a'(x_{t_n}) b(x_{t_n}) + a(x_{t_n}) b'(x_{t_n}) - b(x_{t_n}) \, b'(x_{t_n})^2 - \tfrac{1}{2} b''(x_{t_n}) b(x_{t_n})^2\right] \Delta W_n \Delta t_n \; . \end{aligned} \tag{A.13}$$

For the sake of brevity in the main text, we refer to this method as *Milstein+*. An alternative way to arrive at the approximation

$$I_{(1,0)}(t) \;\rightsquigarrow\; \tfrac{1}{2} \cdot W_t \cdot t \tag{A.14}$$

at the core of this scheme is to look for the *most likely* or the *expected* value of $I_{(1,0)}(t)$ conditional on the discretised path for W_t, i.e. in the filtration generated by the increments ΔW_n. Newton [New91, New94] introduced the idea of *strong asymptotically efficient schemes* following a similar line of reasoning and resulting in the approximation on (A.14). Specifically for the term $I_{(1,0)}(t)$, it so happens that the most likely and the expected value conditional on the discretised path $\{\Delta W_n\}$ are both given by (A.14). Still, the advantage of the Milstein+ scheme over the original Milstein scheme is not that it has a higher convergence order but that it reduces the magnitude coefficient of the leading order error terms.

Applied to equation (2.2), the Milstein+ scheme reads

$$\begin{aligned} V_{t_{n+1}}^{\text{Milstein+}} \;=\; & V_{t_n} + \kappa(\theta - V_{t_n})\Delta t_n + \alpha V_{t_n}^q \Delta Z_n + \tfrac{1}{2}\alpha^2 q V_{t_n}^{2q-1}\left(\Delta Z_n^2 - \Delta t_n\right) + \tfrac{1}{6}\alpha^3 q(2q-1) V_{t_n}^{3q-2} \Delta Z_n^3 \\ & + \tfrac{1}{2}\left[\kappa\alpha\theta q V_{t_n}^{q-1} - \kappa\alpha(q+1) V_{t_n}^q - \tfrac{1}{2}\alpha^3 q\,(3q-1)\, V_{t_n}^{3q-2}\right] \Delta Z_n \Delta t_n \; . \end{aligned} \tag{A.15}$$

A.3 The Milstein scheme for stochastic volatility systems

For the N-dimensional system of stochastic differential equations

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{a}(\boldsymbol{x})\mathrm{d}t + B(\boldsymbol{x})\mathrm{d}\tilde{\boldsymbol{W}} \; , \tag{A.16}$$

with $\boldsymbol{a} \in C^1(\mathbb{R}^N, \mathbb{R}^N)$ and $B \in C^2(\mathbb{R}^N, \mathbb{R}^{N\times N}), \forall i = 1, \ldots, M$ and uncorrelated Brownian motions $\tilde{\boldsymbol{W}} \in \mathbb{R}^N$, the i-th component of the multidimensional Milstein scheme is

$$x_i(t + \Delta t) \;=\; x_i(t) + a_i \Delta t + \sum_{j=1}^{M} b_{ij} \Delta \tilde{W}_j + \underbrace{\sum_{j,k,l=1}^{M} b_{jk}(\partial_{x_j} b_{il}) \tilde{I}_{(k,l)}}_{\text{Milstein term}} \; , \tag{A.17}$$

wherein all of the coefficient functions $a_i(\cdot)$ and $b_{ij}(\cdot)$, etc., are to be evaluated with $\boldsymbol{x}(t)$. We define the double Itô integral $\tilde{I}_{(k,l)}$ as

$$\tilde{I}_{(k,l)} = \int_{s=t}^{t+\Delta t} \int_{u=t}^{s} \mathrm{d}\tilde{W}_k(u)\, \mathrm{d}\tilde{W}_l(s) \ . \tag{A.18}$$

For $k = l$, it simplifies to

$$\tilde{I}_{(k,k)} = \tfrac{1}{2}\left(\Delta\tilde{W}_k^2 - \Delta t\right) \ . \tag{A.19}$$

For the stochastic volatility system (2.1) and (2.2), we use a Cholesky decomposition of the correlated Brownian motions

$$\mathrm{d}W = \rho'\mathrm{d}\tilde{W}_1 + \rho\,\mathrm{d}\tilde{W}_2 \tag{A.20}$$

$$\mathrm{d}Z = \mathrm{d}\tilde{W}_2 \tag{A.21}$$

with $\rho' := \sqrt{1-\rho^2}$ and set $x_1 := \ln S$ and $x_2 := V$ to obtain from (2.1) and (2.2) the coupled stochastic differential equations

$$\begin{pmatrix} \mathrm{d}x_1 \\ \mathrm{d}x_2 \end{pmatrix} = \begin{pmatrix} \mu - \frac{1}{2}x_2^{2p} \\ \kappa(\theta - x_2) \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \rho' x_2^p & \rho\, x_2^p \\ 0 & \alpha x_2^q \end{pmatrix} \begin{pmatrix} \mathrm{d}\tilde{W}_1 \\ \mathrm{d}\tilde{W}_2 \end{pmatrix} \ . \tag{A.22}$$

Since the volatility process x_2 is not influenced directly by the dynamics for x_1, the Milstein scheme for x_2 is given by the standard one-dimensional formula (A.5). For x_1, the fact that $b_{21} = 0$ and $\partial_{x_1} B(\boldsymbol{x}) = 0$ simplifies the calculation of the Milstein term:

$$\begin{aligned} \sum_{j,k,l=1}^{M} b_{jk}(\partial_{x_j} b_{1l})\tilde{I}_{(k,l)} &= b_{22}(\partial_{x_2} b_{11})\tilde{I}_{(2,1)} + b_{22}(\partial_{x_2} b_{12})\tilde{I}_{(2,2)} \\ &= \alpha p x_2^{p+q-1}\left(\rho'\tilde{I}_{(2,1)} + \rho\tilde{I}_{(2,2)}\right) \ . \end{aligned} \tag{A.23}$$

References

[AA00] L. Andersen and J. Andreasen. Volatility Skews and Extensions of the Libor Market Model. *Applied Mathematical Finance*, 7(1):1–32, March 2000.

[Abe04] K.E.S. Abe. Strong Taylor Schemes for Stochastic Volatility. Working paper, 2004. www.maths.ox.ac.uk/~schmitz/project2.htm.

[ABR01] L. Andersen and R. Brotherton-Ratcliffe. Extended Libor Market Models with Stochastic Volatility. Working paper, Gen Re Securities, 2001.

[AP04] L. Andersen and V. Piterbarg. Moment Explosions in Stochastic Volatilty Models. Technical report, Bank of America, 2004. ssrn.com/abstract=559481.

[Bec80] S. Beckers. The constant elasticity of variance model and its implications for option pricing. *Journal of Finance*, XXXV(3):661–673, June 1980.

[BK04] M. Broadie and Ö. Kaya. Exact Simulation of Stochastic Volatility and other Affine Jump Diffusion Processes. Working paper, Columbia University, New York, 2004. www.orie.cornell.edu/~aberndt/FEseminar/papers04/exact_sim_200409.pdf.

[BS73] F. Black and M. Scholes. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, pages 637–654, 1973.

[BS80] M.J. Brennan and E.S. Schwartz. Analyzing convertible bonds. *Journal of Financial and Quantitative Analysis*, 15:907–929, 1980.

[CIR85] J. C. Cox, J. E. Ingersoll, and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–408, 1985.

[CKLS92] C.K. Chan, G.A. Karolyi, F.A. Longstaff, and A.B. Sanders. An empirical comparison of alternate models of the short-term interest rate. *Journal of Finance*, pages 1209–1227, 1992. www.cob.ohio-state.edu/~sanders/ckls.pdf.

[Cox75] J. C. Cox. Notes on option pricing I: Constant elasticity of variance diffusions. Working paper, Stanford University, 1975.

[CR76] J. C. Cox and S. A. Ross. The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3:145–166, March 1976.

[Doo42] J. L. Doob. The Brownian Movement and Stochastic Equations. *The Annals of Mathematics*, 43:351–369, April 1942.

[Dos77] H. Doss. Liens entre équations différentielles stochastiques ordinaires. *Annales de l'Institut Henrie Poincaré. Probabilités et Statistiques*, 13:99–125, 1977.

[Eul68] L. Euler. *Institutiones Calculi Integralis*. 1768.

[GL94] J.G. Gaines and T.J. Lyons. Random Generation of Stochastic Area Integrals. *SIAM Journal on Applied Mathematics*, 54(4):1132–1146, 1994.

[Gla03] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, 2003.

[GY93] H. Geman and M. Yor. Bessel Processes, Asian Options, and Perpetuities. *Mathematical Finance*, 3:349–375, 1993.

[Hes93] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6:327–343, 1993.

[HW88] J. Hull and A. White. An Analysis of the Bias in Option Pricing Caused by a Stochastic Volatility. *Advances in Futures and Options Research*, 3:27–61, 1988.

[HW93] M. Hogan and K. Weintraub. The Lognormal Interest Rate Model and Eurodollar Futures. Discussion paper, Citibank, New York, 1993.

[Jäc02] P. Jäckel. *Monte Carlo methods in finance*. John Wiley and Sons, February 2002.

[Kah04] C. Kahl. Positive numerical integration of stochastic differential equations. Master's thesis, Bergische Universität Wuppertal, 2004. www.math.uni-wuppertal.de/~kahl/publications/DT.pdf.

[KJ05] C. Kahl and P. Jäckel. Not-so-complex logarithms in the Heston model. *Wilmott*, September, September 2005.

[KP99] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992, 1995, 1999.

[KS91] I. Karatzas and S. E. Shreve. *Brownian motion and Stochastic Calculus*. Springer, 1991.

[KS05] C. Kahl and H. Schurz. Balanced Milstein Methods for Ordinary SDEs. Technical report, Department of Mathematics, Southern Illinois University, 2005.

[KT81] S. Karlin and M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, 1981.

[Lév51] P. Lévy. Wiener's random function, and other Laplacian random functions. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 171–187, Berkeley and Los Angeles, 1951. University of California Press.

[MAP76] G. Marsaglia, K. Anantharayanan, and N.J. Paul. Improvements on fast methods for generating normal random variables. *Information Processing Letters*, 5:27–30, 1976.

[Mar55] G. Maruyama. Continuous Markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo*, 4:48–90, 1955.

[Mil74] G. N. Milshtein. Approximate integration of stochastic differential equations. *Theory of Probability and Applications*, 19:557–562, 1974.

[MMB64] G. Marsaglia, M.D. MacLaren, and T.A. Bray. A fast procedure for generating normal random variables. *Communications of the ACM*, 5:27–30, 1964.

[MN98] M. Matsumoto and T. Nishimura. Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Trans. on Modeling and Computer Simulation*, 8(1):3–30, January 1998.

[MPS98] G. N. Milshtein, E. Platen, and H. Schurz. Balanced implicit methods for stiff stochastic systems. *SIAM*, 38(3):1010–1019, 1998.

[New91] N. J. Newton. Asymptotically efficient Runge-Kutta methods for a class of Ito and Stratonovich equations. *SIAM Journal of Applied Mathematics*, 51:542–567, 1991.

[New94] N. J. Newton. Variance reduction for simulated diffusions. *Siam Journal of Applied Mathematics*, 54:1780–1805, 1994.

[PT85] E. Pardoux and M. Talay. Discretization and simulation of stochastic differential equations. *Acta Applicandae Mathematica*, 3:23–47, 1985.

[Sch96] H. Schurz. Numerical Regularization for SDE's: Construction of nonnegative solutions. *Dynamical Systems and Applications*, 5:323–352, 1996.

[Sco87] L. Scott. Option Pricing When the Variance Changes Randomly: Theory, Estimation and An Application. *Journal of Financial and Quantitative Analysis*, 22:419–438, December 1987.

[SS91] E. M. Stein and J. C. Stein. Stock Price Distribution with Stochastic Volatility : An Analytic Approach. *Review of Financial Studies*, 4:727–752, 1991.

[SS94] K. Sandmann and D. Sondermann. On the stability of log-normal interest rate models and the pricing of Eurodollar futures. Discussion paper, Dept. of Statistics, Faculty of Economics, SFB 303, Universität Bonn, June 1994. ftp://ftp.wipol.uni-bonn.de/pub/RePEc/bon/bonsfb/bonsfb263.pdf.

[SS97a] K. Sandmann and D. Sondermann. A Note on the Stability of Lognormal Interest Rate Models and the Pricing of Eurodollar Futures. *Mathematical Finance*, 7(2):119, April 1997.

[SS97b] K. Sandmann and D. Sondermann. Log-Normal Interest Rate Models: Stability and Methodology. Discussion paper, Dept. of Statistics, Faculty of Economics, SFB 303, Universität Bonn, January 1997. ftp://ftp.wipol.uni-bonn.de/pub/RePEc/bon/bonsfb/bonsfb398.pdf.

[SZ99] R. Schöbel and J. Zhu. Stochastic Volatility With an Ornstein Uhlenbeck Process: An Extension. *European Finance Review*, 3:23–46, 1999. ssrn.com/abstract=100831.

[UO30] G.E. Uhlenbeck and L.S. Ornstein. On the theory of Brownian motion. *Physics Review*, 36:823–841, 1930.

[Wig87] J. Wiggins. Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics*, 19:351–372, 1987.

[WZ65] E. Wong and M. Zakai. On the convergence of ordinary integrals to stochastic integrals. *Ann. Math. Stat.*, 36:1560–1564, 1965.